

University of Dundee

Quantification of Differential Transcription Factor Activity and Multiomics-Based Classification into Activators and Repressors

Berest, Ivan; Arnold, Christian; Reyes-Palomares, Armando; Palla, Giovanni; Rasmussen, Kasper Dindler; Giles, Holly

Published in:
Cell Reports

DOI:
[10.1016/j.celrep.2019.10.106](https://doi.org/10.1016/j.celrep.2019.10.106)

Publication date:
2019

Licence:
CC BY

Document Version
Publisher's PDF, also known as Version of record

[Link to publication in Discovery Research Portal](#)

Citation for published version (APA):

Berest, I., Arnold, C., Reyes-Palomares, A., Palla, G., Rasmussen, K. D., Giles, H., Bruch, P-M., Huber, W., Dietrich, S., Helin, K., & Zaugg, J. B. (2019). Quantification of Differential Transcription Factor Activity and Multiomics-Based Classification into Activators and Repressors: diffTF. *Cell Reports*, 29(10), 3147-3159.e12. <https://doi.org/10.1016/j.celrep.2019.10.106>

General rights

Copyright and moral rights for the publications made accessible in Discovery Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from Discovery Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain.
- You may freely distribute the URL identifying the publication in the public portal.

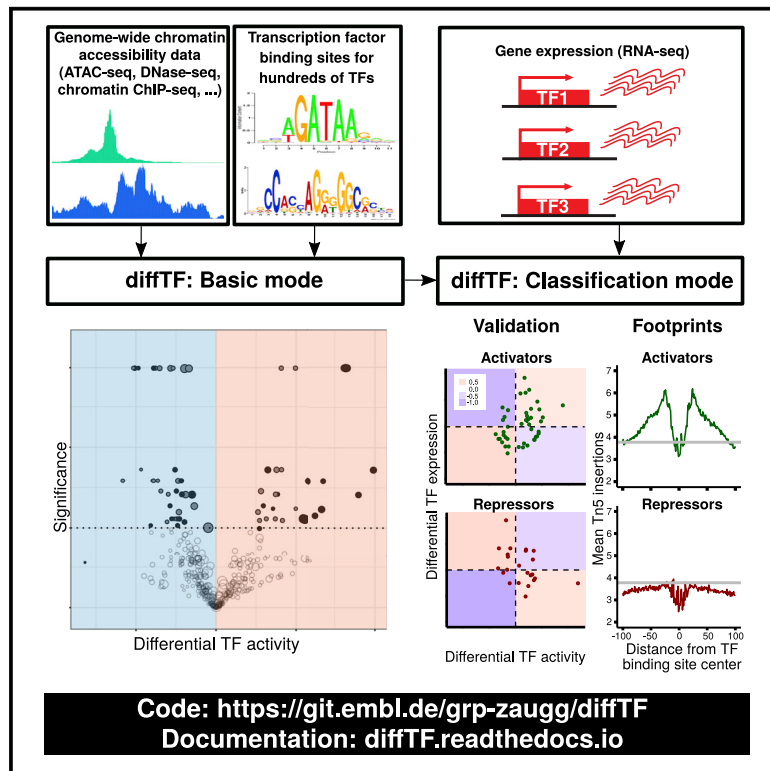
Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Cell Reports

Quantification of Differential Transcription Factor Activity and Multiomics-Based Classification into Activators and Repressors: diffTF

Graphical Abstract



Authors

Ivan Berest, Christian Arnold, Armando Reyes-Palomares, ..., Sascha Dietrich, Kristian Helin, Judith B. Zaugg

Correspondence

judith.zaugg@embl.de

In Brief

Berest et al. present a computational tool (diffTF) to estimate differential TF activity and classify TFs into activators or repressors. It requires active chromatin data (accessibility/ChIP-seq) and integrates with RNA-seq for classification. The authors apply it to two case studies (CLL and hematopoietic differentiation) and validate their predictions experimentally.

Highlights

- diffTF estimates differential TF activity based on chromatin data (basic mode)
- Integration with RNA-seq identifies activator/repressor TFs (classification mode)
- We applied diffTF to subtypes of CLL and hematopoietic differentiation
- The TF classification was experimentally validated in both case studies



Quantification of Differential Transcription Factor Activity and Multiomics-Based Classification into Activators and Repressors: diffTF

Ivan Berest,^{1,9,10} Christian Arnold,^{1,10} Armando Reyes-Palomares,¹ Giovanni Palla,¹ Kasper Dindler Rasmussen,^{2,5} Holly Giles,^{3,4,9} Peter-Martin Bruch,³ Wolfgang Huber,^{1,4,6} Sascha Dietrich,³ Kristian Helin,^{5,7,8} and Judith B. Zaugg^{1,4,6,11,*}

¹Structural and Computational Biology Unit, European Molecular Biology Laboratory, Heidelberg, Germany

²School of Life Sciences, University of Dundee, Dundee, UK

³Heidelberg University Hospital, Heidelberg, Germany

⁴Genome Biology Unit, European Molecular Biology Laboratory, Heidelberg, Germany

⁵Biotech Research and Innovation Centre (BRIC), University of Copenhagen, Copenhagen, Denmark

⁶European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, Cambridge, UK

⁷Novo Nordisk Foundation Center for Stem Cell Biology, Copenhagen, Denmark

⁸Cell Biology Program and Center for Epigenetics Research, Memorial Sloan Kettering Cancer Center, New York, NY, USA

⁹Collaboration for joint PhD degree between EMBL and Heidelberg University, Faculty of Biosciences, Heidelberg, Germany

¹⁰These authors contributed equally

¹¹Lead Contact

*Correspondence: judith.zaugg@embl.de
<https://doi.org/10.1016/j.celrep.2019.10.106>

SUMMARY

Transcription factors (TFs) regulate many cellular processes and can therefore serve as readouts of the signaling and regulatory state. Yet for many TFs, the mode of action—repressing or activating transcription of target genes—is unclear. Here, we present diffTF (<https://git.embl.de/grp-zaugg/diffTF>) to calculate differential TF activity (basic mode) and classify TFs into putative transcriptional activators or repressors (classification mode). In basic mode, it combines genome-wide chromatin accessibility/activity with putative TF binding sites that, in classification mode, are integrated with RNA-seq. We apply diffTF to compare (1) mutated and unmutated chronic lymphocytic leukemia patients and (2) two hematopoietic progenitor cell types. In both datasets, diffTF recovers most known biology and finds many previously unreported TFs. It classifies almost 40% of TFs based on their mode of action, which we validate experimentally. Overall, we demonstrate that diffTF recovers known biology, identifies less well-characterized TFs, and classifies TFs into transcriptional activators or repressors.

INTRODUCTION

Transcription factors (TFs) coordinate dynamic responses to intra- and extracellular stimuli and regulate a multitude of biological processes. Because many signaling cascades activate a specific set of TFs, observing a change in overall TF activity can serve as a proxy for activity of signaling pathways (Kim et al., 2007). The activity of a TF is often cell type specific and de-

pends on cofactors, binding partners, and local chromatin environment (Whyte et al., 2013). Adding to this complexity, the mode of action of many TFs can vary, making them act as transcriptional activators and repressors, depending on the context (Han et al., 2018). Thus, to correctly interpret the downstream effects of a change in TF abundance, it is important to understand its global mode of action within the context of the study.

TFs are typically lowly abundant proteins and therefore difficult to detect in proteomics experiments (Kim et al., 2007; Teng et al., 2008). If they are detected, their abundance and activity do not necessarily correspond, because TFs are highly regulated at the posttranslational level. However, chromatin immunoprecipitation followed by sequencing (ChIP-seq), which is the gold-standard technique for measuring genomic TF binding events, only provides information for one TF at a time and does not detect global changes in TF binding activity unless specific experimental normalization methods are used (e.g., spike-ins) (Bonhoure et al., 2014). In addition, neither proteomics nor ChIP-seq experiments measure a TF's mode of action. Activity and mode of action can be measured by luciferase assays; however, they are typically done for a specific TF at a specific location and therefore have fairly low throughput (Komatsu et al., 2010; Liu et al., 2011). Databases like TRRUST (Han et al., 2018) are collecting annotations of regulation modes of TFs based on the literature and provide a comprehensive resource for well-studied TF-target interactions. However, for most TFs, there is contradictory evidence about their molecular mode of action. In summary, a general framework for determining differential activity of TFs between conditions and classifying TFs into transcriptional activators and repressors in a cell-type- and condition-specific manner is currently lacking.

To move toward closing these gaps, we developed diffTF, a tool to estimate global changes in TF activities and to classify them into activators and repressors based on the integration of genome-wide chromatin accessibility or histone mark ChIP-seq



data with predicted TF binding sites and RNA sequencing (RNA-seq) data. We have extensively tested and validated our approach in two case studies: case study I, comparing two patient cohorts, each with a large number of heterogeneous samples, and case study II, comparing two cell types along a differentiation trajectory, each with a small number of genetically identical and thus homogeneous replicates. For case study I, we obtained publicly available ATAC-seq data of a cohort of >50 patients suffering from chronic lymphocytic leukemia (CLL) (Rendeiro et al., 2016). For case study II, we performed ATAC-seq and RNA-seq on murine multipotent progenitors (MPPs) versus granulocyte-macrophage progenitors (GMPs) in quadruplicate (Rasmussen et al., 2019). In both case studies, diffTF recovered the known biology, and it was able to classify 39% and 63% of all expressed TFs based on their mode of action in case study I and II, respectively. We validated the classification experimentally by comparing differential expression and differential activity of TFs in each cell type upon drug stimulation (case study I) and genetic perturbation (case study II).

RESULTS

Conceptual Derivation of TF Activity and Classification of TFs into Transcriptional Activators and Repressors

We define TF motif activity, or TF activity for short, as the effect of a TF on the state of chromatin as measured by chromatin accessibility or active chromatin marks (i.e., ATAC-seq, DNase sequencing [DNase-seq], or histone H3 lysine 27 acetylation [H3K27ac] ChIP-seq). This definition is based on our earlier work, in which we showed that genetic variants affecting H3K27ac signal across individuals can be explained by disruptions of TF motifs whenever the variant overlaps with a TF binding site (TFBSs) (Grubert et al., 2015). This suggests that TFs play a causal role in mediating the effect of the DNA variant onto chromatin marks (Liu et al., 2015). By reversing this argument, we propose to use the aggregate changes in chromatin accessibility near putative binding sites of a TF as a readout for its potential effect on chromatin. A similar concept has been proposed in other tools that estimate TF activity based on ATAC-seq or DHS data (Baek et al., 2017; Schep et al., 2017).

Because little is known about the functional mode of action of TFs, and indeed, most TFs have been reported to act as both activator and repressor, depending on the study (Han et al., 2018), we wanted to devise a data-driven approach to classify TFs into potential activators and repressors. The classification framework is based on the following assumptions: (1) Increasing the abundance of an activator TF will increase the average accessibility at the regulatory elements controlled by the TF, which will lead to upregulation of its target genes. (2) Increasing the abundance of a repressor TF will decrease the average accessibility at the regulatory elements controlled by the TF, which will lead to a downregulation of its target genes.

Although the assumption for activators is well accepted, it is less straightforward for repressors: on the one hand, the binding of the TF increases the accessibility locally; on the other hand, repression is globally associated with closed chromatin. Therefore, we test these assumptions in the context of case study I (see the next section).

These ideas are implemented in diffTF to globally assess differential TF activity between two conditions (basic mode; Figure 1A; Figure S1A) and to classify TFs into activators and repressors (classification mode; Figure 1B). Basic mode is based on genome-wide accessibility or ChIP-seq (of active chromatin marks) data and a list of putative TFBSs that can be obtained by *in silico* predictions (e.g., using PWMScan; Ambrosini et al., 2018) or from a collection of ChIP-seq data (e.g., ReMap; Griffon et al., 2015). It outputs differential TF activity and a p value for each TF. Briefly, diffTF calculates the log2 fold change in accessibility (or histone ChIP-seq) for each TFBS and summarizes them per TF while normalizing for GC content (Figure S1B; see STAR Methods). The significance is assessed using either an empirical test, based on TF activities from permuted sample labels, or an analytical procedure, which provides a p value based on t statistic and estimated variance (see STAR Methods; Figures S2D and S2E). For classification mode, diffTF additionally requires RNA-seq data to estimate the abundance of each TF by its expression level (see STAR Methods). It outputs a classification into putative activators, repressors, or undetermined for each TF. To do so, it compares the distribution of correlations for peaks with putative binding sites (foreground) against all other peaks (background) and classifies TFs depending on whether the correlations of putative targets are significantly more positive than (activator), more negative than (repressor), or indistinguishable from (undetermined) the background (Figure 1B).

Case Study I: Differential TF Activity in a Heterogeneous ATAC-Seq Dataset in CLL

To assess the technical robustness of diffTF and its power to recover a relevant biological signal, we applied it to a well-studied biological system, which we found in a study that comprises a large ATAC-seq dataset comparing different subtypes of the extensively studied cancer chronic lymphocytic leukemia (CLL) (Rendeiro et al., 2016).

Chronic lymphocytic leukemia (CLL) is a frequent cancer in the Western world, particularly among the elderly. It falls into two major subtypes, which are defined based on the mutation status of the IgHV locus (mutated CLL [M-CLL] and unmutated CLL [U-CLL]). In M-CLL, B cells go through normal affinity maturation with the aid of T helper cells and undergo multiple rounds of IgHV somatic hypermutation to produce high-affinity B cell receptors (BCRs). In contrast, U-CLL B cells reach their affinity maturation in an unregulated manner without T helper cells (Chiorazzi and Ferrarini, 2011). Overall, this leads to worse clinical outcomes, shorter survival, and higher relapse frequency (Furman et al., 2014).

The dataset comprises of ATAC-seq data for 56 CLL patients (88 samples) that are stratified by the mutation status of the IgHV locus (34 U-CLL, 50 M-CLL, and 4 unclassified; see Table S3). After data processing and quality control, 25 U-CLL and 27 M-CLL samples remained (Figure S3; see STAR Methods). Applying diffTF in basic mode revealed 68 differentially active TFs between the two subtypes of CLL (false discovery rate [FDR] < 10%; Figure 2A; Table S1).

Because TF binding site predictions, on which diffTF relies, are inherently noisy and may result in many false-positive

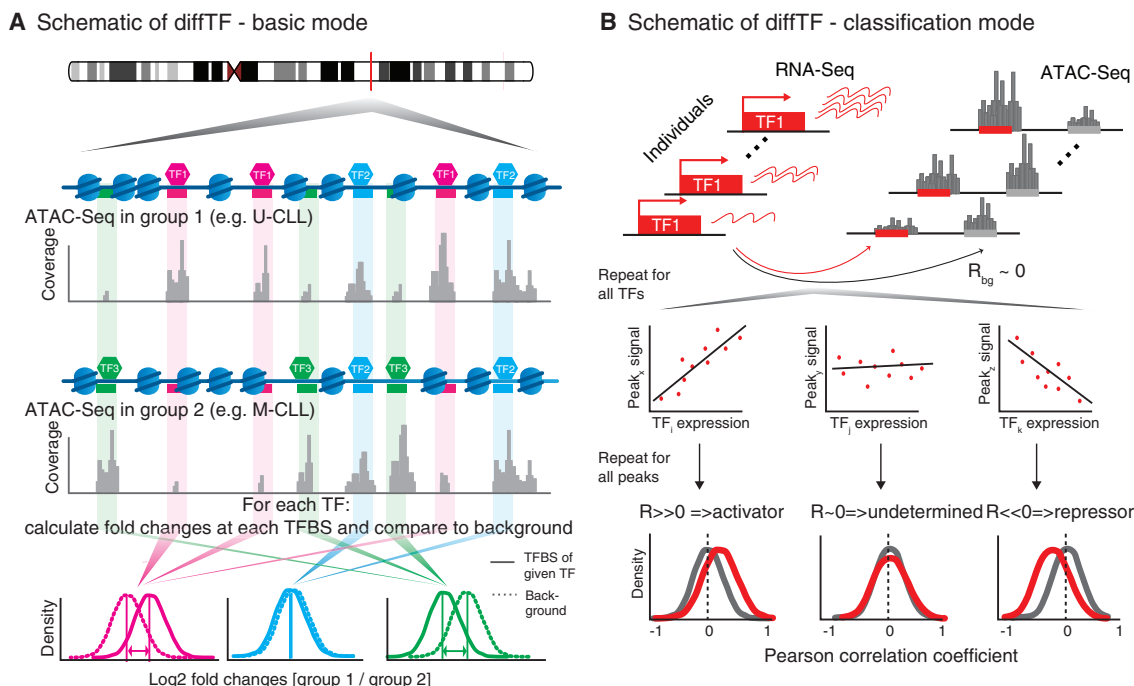


Figure 1. Schematic Representation of the diffTF Workflow

(A) Simplified workflow illustrating the diffTF basic mode: For each binding site of a given TF (TFBS), the fold change between the two conditions is computed, followed by comparing their distribution to a background set of fold changes obtained from GC content-matched loci that do not contain the putative TFBS. The difference in the distributions is visualized in a volcano plot, where the y axis indicates statistical significance and the x axis indicates the effect size. (For a detailed workflow, see Figure S1A.)

(B) Schematic representation of the diffTF classification mode: TF expression levels are correlated with the accessibility of their target sites. If correlations with its target sites are more positive than with the background distribution (non-target sites), the TF is classified as putative activator; if they are more negative than with the background, it is classified as putative repressor; and if they are indistinguishable from the background, it is classified as undetermined (see STAR Methods).

predictions when compared with ChIP-seq experiments (Landt et al., 2012), we wanted to ensure the robustness of diffTF before focusing on the biological interpretation. To do so, we compared the results of a diffTF run on all predicted binding sites (see STAR Methods) to those of a run on ChIP-seq-validated sites only and found that they strongly correlate (Pearson's $R = 0.84$; Figure 2B; Figures S4A and S4B). Similarly, we found diffTF is highly robust with respect to other parameters for TF binding site predictions, such as nucleotide composition of the predicted binding sites for PWMScan (Figures 4C and 4D), p value thresholds in PWMScan to predict TFBS ($1e-5$ versus either $1e-6$ or $5e-5$; Figures 2C and 2D, $R = 0.62$ and 0.87 , respectively), and motif database (JASPAR versus HOCOMOCO; Figure 2E, $R = 0.69$). Additional robustness tests such as the impact of different motif extension sizes are described in STAR Methods and Figures 4E and 4F. Overall, these tests demonstrate the robustness of diffTF with respect to the high noise inherent to TFBS predictions. The reasons for the robustness are likely that diffTF aggregates signals across thousands of TFBSs for each TF and that false-positive predictions increase the noise without eliminating the signal.

We hypothesized that aggregating signals across many TFBS per TF may even allow diffTF to detect differences in experiments with little biological signal. To test this, we removed high-signal regions (i.e., differentially accessible peaks at 5% FDR; see STAR Methods) and found that the resulting differential

TF activities were similar to those obtained from a run on the full set ($R = 0.89$), thus demonstrates the power of diffTF to capture the differential TF activities in low-signal settings (Figure 2F). In agreement with this, we found that diffTF provides highly congruent results across a range of sample sizes and sequencing depths (Figure S4G). Generally, the number of samples appeared more important than the read depth, with even 1 million to 5 million processed reads providing consistent results when using all samples (see STAR Methods). Although these results are dataset specific, they provide guidelines for the applicability of diffTF and are in line with single-cell ATAC-seq data analyses that show robustness to low coverage for genome-wide summary statistics (Mezger et al., 2018).

In summary, these results establish the robustness of diffTF in quantifying differences in TF activities and demonstrate that aggregating signals across all binding sites is a powerful and sensitive approach to overcome limitations such as low coverage and little underlying biological signal.

diffTF Proposes Many Previously Unreported TF Candidates that Distinguish U-CLL from M-CLL

We next focused on the biological interpretation of the differentially active TFs between M-CLL and U-CLL (FDR 10%; Figure 2A). About 44% of them have previously been associated with CLL, and most of them (80%) agree with the reported

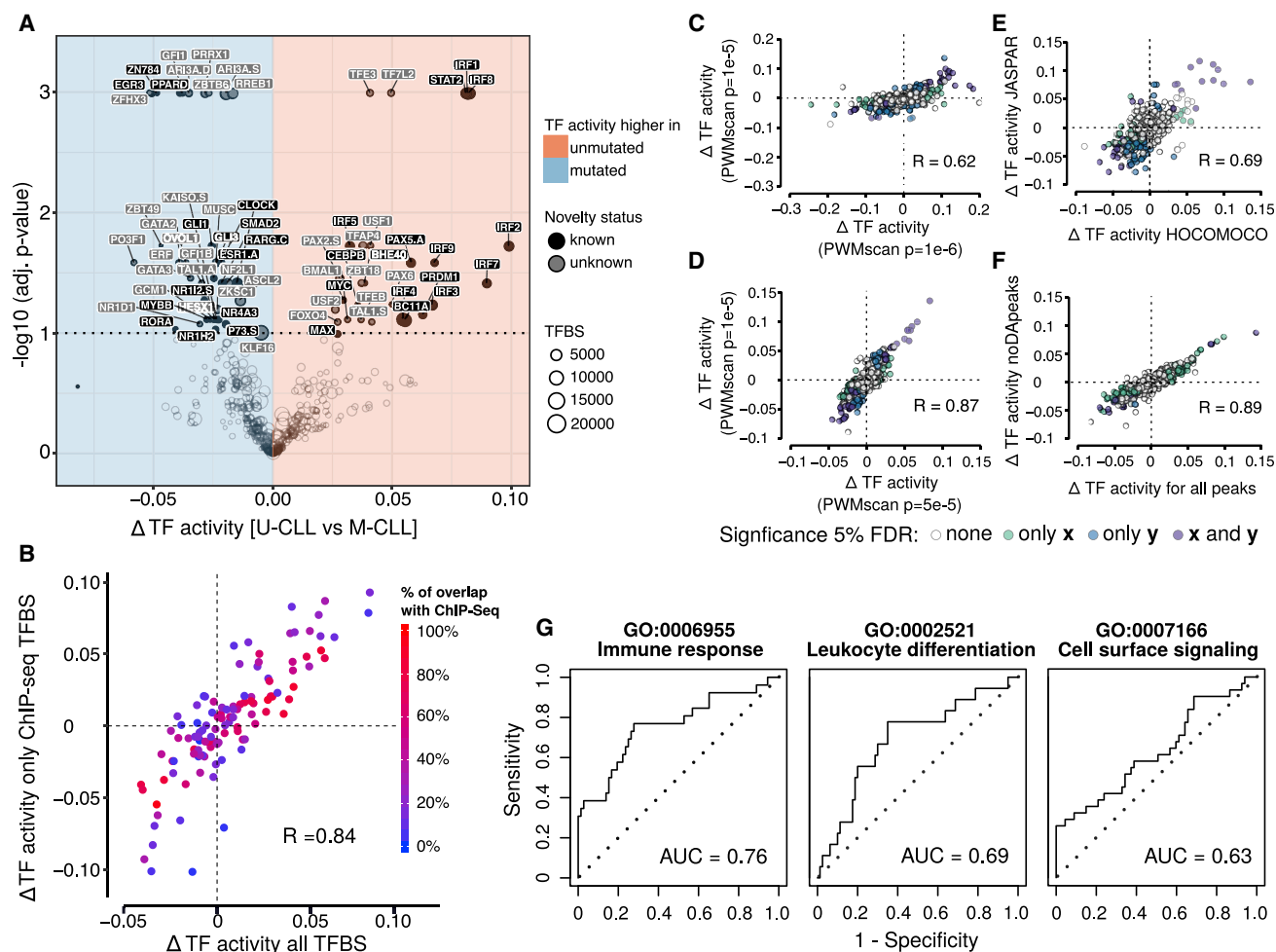


Figure 2. Technical Robustness and Biological Validation of Results from diffTF in Basic Mode Applied to Case Study I

(A) Volcano plot of differential TF activity between U-CLL (n = 27 biological replicates) and M-CLL (n = 25 biological replicates). Significance threshold (10% FDR) is indicated with a dotted line. TFBS, number of predicted TFBSs. p values are obtained through diffTF using the empirical approach and adjusted by the Benjamini-Hochberg procedure (y axis).

(B–F) Assessment of technical robustness. (B) Comparison of all predicted TFBSs and TFBSs experimentally validated by ChIP-seq data from ReMap (see also Figures S4A and S4B). For each TF (n = 640), the percentage of TFBSs that overlap ChIP-seq data is indicated from blue (0%) to red (100%). (C and D) Comparisons between p value thresholds in PWMScan to predict TFBS (n = 628 TFs): (C) standard (1e–5) versus stringent (1e–6) and (D) standard (1e–5) versus relaxed (5e–5). (E) Comparisons between different motif databases (HOCOMOCO v.10 versus JASPAR 2016; n = 412 TFs). (F) Comparisons between different peak sets (full consensus [allPeaks] versus non-differentially accessible peaks [noDApeaks]; n = 640 TFs). For (C)–(F), R indicates Pearson correlation and TFs are colored by diffTF significance (5% FDR) in the compared analyses (white, not significant; light green or blue, significant for the x axis or y axis only; purple, significant in both).

(G) Receiver-operator characteristic (ROC) curves for three selected Gene Ontology (GO) terms with high area under the curve (AUC) based on all differentially active TFs (FDR < 10%) between U-CLL and M-CLL. See Table S2 for the full list of significant GO terms.

condition (i.e., M-CLL or U-CLL; Table S2). Overall, we found that the differential TFs recapitulate many known biological differences between M-CLL and U-CLL, such as leukocyte differentiation, immune response, and cell-surface signaling that includes BCR signaling, all known to be increased in U-CLL (Figure 2G; Table S2; see STAR Methods).

TFBS prediction is inherently based on TF motifs, which can be similar across related TFs. We therefore grouped TFs into TF motif families using the position weight matrix (PWM) clustering tool from RSAT (Castro-Mondragon et al., 2017; clusters available at <https://bit.ly/2J9Taak>). Most significant TFs fell

into ten groups (Figure S5A). The most active groups of TFs in U-CLL are the IRF family and STAT2 (Figure 2A), both of which have been associated with disease onset and progression and harbor several CLL susceptibility loci (Arvaniti et al., 2011; Have-lange et al., 2011; Slager et al., 2013). They are followed by a group of TFs that include c-MYC, a cell proliferation factor that is highly abundant in U-CLL (Landau et al., 2015; Yeomans et al., 2016). For M-CLL, we identified TFs that suggest normal functionality of B cells through the classical BCR, nuclear factor κ B (NF- κ B), and Wnt signaling pathways (Table S2). The most active TF family for M-CLL were the ROR factors (Figure S5A),

which coactivate NF- κ B-dependent survival signaling in CLL (Minami et al., 2010) and the GATA family, known to prime hematopoietic stem cells (HSCs) toward the lymphoid lineage and to increase self-renewal of the stem cells in CLL (Kikushige et al., 2011). We also found the EGR family, which are implicated in aberrantly hypomethylated CpG sites in CLL (Oakes et al., 2016); Gli1, which is part of the Hedgehog signaling pathway that regulates apoptosis, thereby supporting survival of M-CLL cells (Kern et al., 2015); and PPARG, which has recently been linked to M-CLL through its effect on metabolic pathways in cancer cells (Li et al., 2017).

Among the factors that have not previously been associated with CLL we found several TFs significantly higher in U-CLL that are involved in the regulation of the circadian clock (i.e., BMAL1, CLOCK, and NR1D1), which has recently been proposed as a hallmark of cancer (El-Athman and Relógio, 2018). Other TFs include members of the basic helix-loop-helix family, such as BHE40, a regulator of mitotic division (D'Annibale et al., 2014) and essential for the development of B1-a cells (Kreslavsky et al., 2017), and TFAP4, TFE3, and TFEB, for which there are known cases of gene fusions in renal cell carcinoma (Kauffman et al., 2014). TFs more active in M-CLL were associated with pathways relevant for cancer cells and B cells, such as escape from apoptosis (ZN784) (Kasim et al., 2017), regulation of cell-cycle progression (ZBTB6) (Chevrier et al., 2014), and selection of B cells and promotion of fetal B lymphopoiesis (ARID3A) (Zhou et al., 2015). GFI1 family members are less active in U-CLL, and their expression and activation might influence and decrease rates of apoptosis in B cells (Coscia et al., 2011).

In summary, these results show that diffTF recapitulates much of the known biology of the two subtypes of CLL and identifies several TFs that have not previously been associated with CLL and that are likely implicated in the subtypes of the disease.

Determination of the Molecular Function of TFs: Transcriptional Repressors and Activators

To gain more mechanistic insights into some regulatory differences between U-CLL and M-CLL, we applied diffTF in classification mode. In addition, we used these data to test the assumption posed in the beginning that repressors and activators have an opposing effect on chromatin accessibility at their target sites, which underlies our classification framework.

To understand the effect of repressors and activators on chromatin accessibility, we compared the accessibility footprint (Tn5 insertion sites) of a well-known repressor (REST) and a well-known activator (STAT2) that are expressed in our cell type. For REST, we found an increase in accessibility at the borders of its motif, which likely reflects accessibility induced by the binding of the TF. Importantly, however, and in line with our assumptions, we found that the accessibility drops below the genome-wide average within 10 bp of the motif center (Figure 3A, bottom). This is in contrast to STAT2, which showed increased chromatin accessibility outside its core binding site, slowly dropping to the genome-wide average over a distance > 100 bp from the center of the motif and likely representing the effect of the TF on opening the surrounding chromatin (Figure 3A, top). Thus, even though there is an increase in accessibility for the repressor

at the immediate binding site, the surrounding chromatin is highly compact, whereas it is open for the activator. This provides evidence that the assumption underlying our TF classification approach is biologically reasonable.

Applying diffTF in classification mode to the CLL dataset, for which RNA-seq data were available for eight individuals (after QC; see STAR Methods), we were able to classify 39% of the expressed TFs (146 of 370) as either putative activators or repressors (STAR Methods; Figure S5B). In Figure 3B, we show an example of the foreground and background distributions for an activator (IRF2), a repressor (PAX5), and an unclassified TF (SP4). To visually assess these distributions for all TFs, we subtracted the background from the foreground distribution and displayed it as a heatmap (Figure 3C). This illustrates the shifts we observe for putative activators and repressors yet shows that a large fraction of TFs have similar foreground and background distributions and are therefore classified as undetermined. Our classification is based on variation across individuals; thus, we will not be able to classify TFs with low expression variation. Other reasons for undetermined classifications include post-translational regulation, such that RNA expression does not capture the abundance of a TF's active form, or bifunctional TFs that act as activator and repressor in the same cell type.

Footprint analyses of all differentially active TFs followed by principal-component analysis (PCA) revealed four major footprint classes that were distinguishable by the accessibility at the binding site and flanks (as captured by PC1 and PC2, respectively; Figure 3D). Most TFs classified as repressors fell into classes III and IV (37 of 57), which are characterized by low flank accessibility (low PC2), whereas most putative activators fell into classes I and II (53 of 89), which are characterized by high flank accessibility (high PC2). This typifies class III and IV, which included PAX5 (Figure 3E, left), as repressors and class I and II, which included STAT2 (Figure 3E, right), as activators. When plotting the footprints for all predicted activators and all predicted repressors, they overall recapitulate their predicted molecular mode of action with respect to opening or closing chromatin in the flanking regions of their binding sites (Figure 3F).

Validation of the Activator/Repressor Classification and Their Characterization

When we compared our TF classifications to the mode of action that was curated by the TRRUST database, we found that most TFs have been classified almost equally often as both activators and repressors (Figure S5C). Therefore, we had to find another way to validate our predictions. To do so, we devised the following testing framework: if our classifications were correct, we would expect that increasing the abundance of a TF results in increased motif accessibility for activators and decreased motif accessibility for repressors. Therefore, differential TF activity and differential expression should be positively correlated for activators and negatively correlated for repressors. We assessed this expectation within the Rendeiro et al. (2016) data: We calculated differential expression between U-CLL and M-CLL for all TFs stratified into activators, repressors, and undetermined and compared this to their differential TF activity. The results agreed with our expectations of a positive correlation for activators (Spearman's $\rho = 0.22$, $p = 0.033$), a negative correlation for

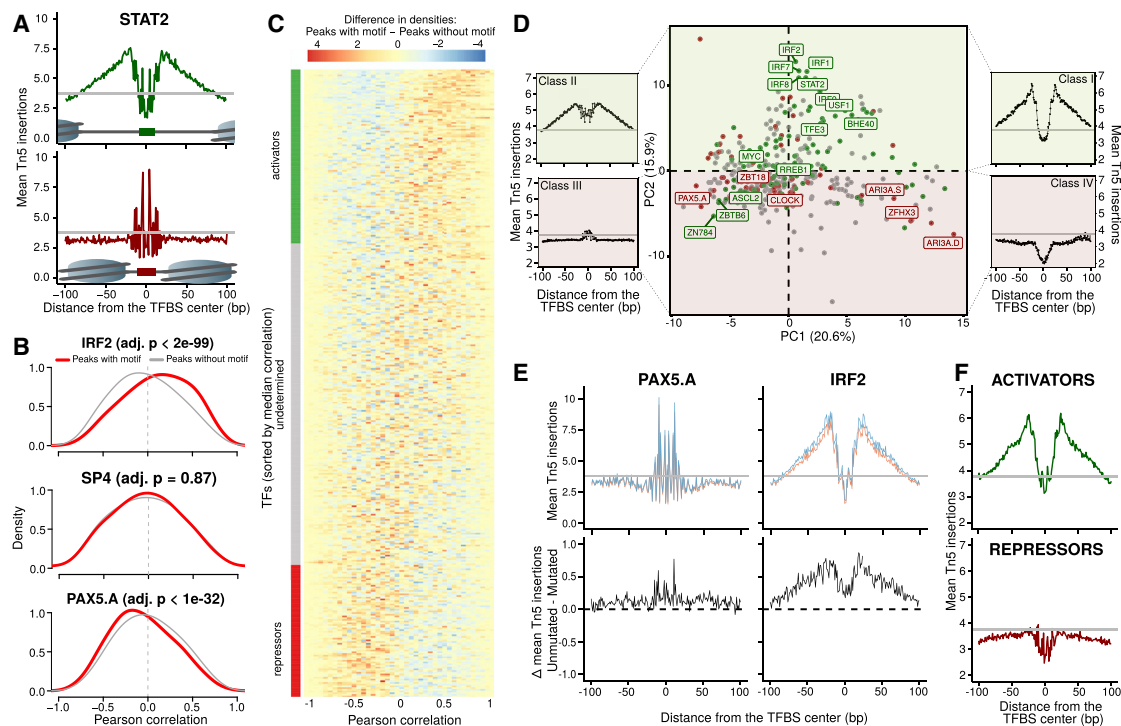


Figure 3. Characterization of Results from diffTF in Classification Mode Applied to Case Study I

(A) Exemplary footprints (mean Tn5 insertions centered at TFBS) for a well-known activator (STAT2, top) and a well-known repressor (REST, bottom). Tn5 insertions were normalized to the library size and numbers of samples between U-CLL and M-CLL. The genome-wide average of insertions within accessible chromatin is shown as a solid gray line.

(B) Foreground (red) and background (gray) distributions of Pearson correlations between RNA expression and signal at ATAC-seq peaks are shown for a putative activator (IRF2, $n = 4,362$ TFBS, top), a repressor (PAX5, $n = 4,438$ TFBS, bottom), and an undetermined TF (SP4, $n = 15,938$ TFBS, middle). The p values are given for Wilcoxon rank-sum test and adjusted using the Benjamini-Hochberg procedure.

(C) Difference between foreground and background distributions for all TFs are displayed as a heatmap. Each horizontal line represents the subtraction of the binned foreground minus the binned background correlation distributions (40 bins) for one TF. TFs are sorted from strongest predicted activator to repressor.

(D) Scatterplot comparing PC1 and PC2 from a PCA of the footprints of all expressed TFs ($n = 370$). The insets (classes I–IV) display the average footprint across TFs in that quadrant.

(E) Footprint analysis for one of the strongest putative activators (IRF2, right) and one of the strongest putative repressors (PAX5.A, left). Footprints are shown separately for M-CLL (blue) and U-CLL (orange) based on the normalized number of Tn5 insertions (top) and as the difference (U-CLL minus M-CLL, bottom).

(F) Average footprints for all significant ($FDR < 10\%$) activators (top) and repressors (bottom) are shown.

See (A) for axis descriptions. The data presented in this figure are based on 27 and 25 biological replicates for U-CLL and M-CLL, respectively.

repressors ($\rho = -0.37$, $p = 0.01$), and no correlation for undetermined TFs ($\rho = -0.04$, $p = 0.51$; Figure 4A, top). Because one could argue that this is a bit circular (TF classifications are based on the same data as the validations), we next assessed the classifications in an independent experiment. To do so, we obtained mononuclear cells from four CLL patients of an independent patient cohort (H.G., unpublished data) and profiled RNA-seq and ATAC-seq with and without treating them with ibrutinib (see STAR Methods). We then calculated TF activity and differential TF expression before and after treatment while using the same classification from case study I to stratify the TFs into putative activators and repressors. We found that 46 of 68 TFs changed in the predicted direction, i.e., a change in abundance of a predicted activator was positively correlated with its change in TF activity ($\rho = 0.29$, $p = 0.054$), while it was negatively correlated for predicted repressors ($\rho = -0.40$, $p = 0.05$) (Figure 4A, bottom). Overall, these results demonstrate that the transcriptional upregulation of TFs upon

ibrutinib treatment resulted in a TF activity change that was consistent with our prediction in 68% of the differentially active TFs.

To further characterize our putative repressors and activators, we overlapped their binding sites with chromatin states for primary B cells obtained from the Epigenomic Roadmap (Kundaje et al., 2015). We found that our predicted activators and repressors preferentially locate to active and repressive states, respectively (Figure 4B; Figure S5D). This corroborates our molecular mode of action predictions and is in line with the observation that repressor footprints tend to have less accessible chromatin in their flanking regions than activators (see Figure 3A).

Finally, we verified that our measure of chromatin accessibility-based TF activity reflects a TF's effect on its target genes by comparing the direction of expression change of target genes of a given TF with its change in activity as calculated by diffTF. The direction agreed in 29 of the 40 tested TFs (Figure 4C; see STAR Methods). Importantly, this relationship held true

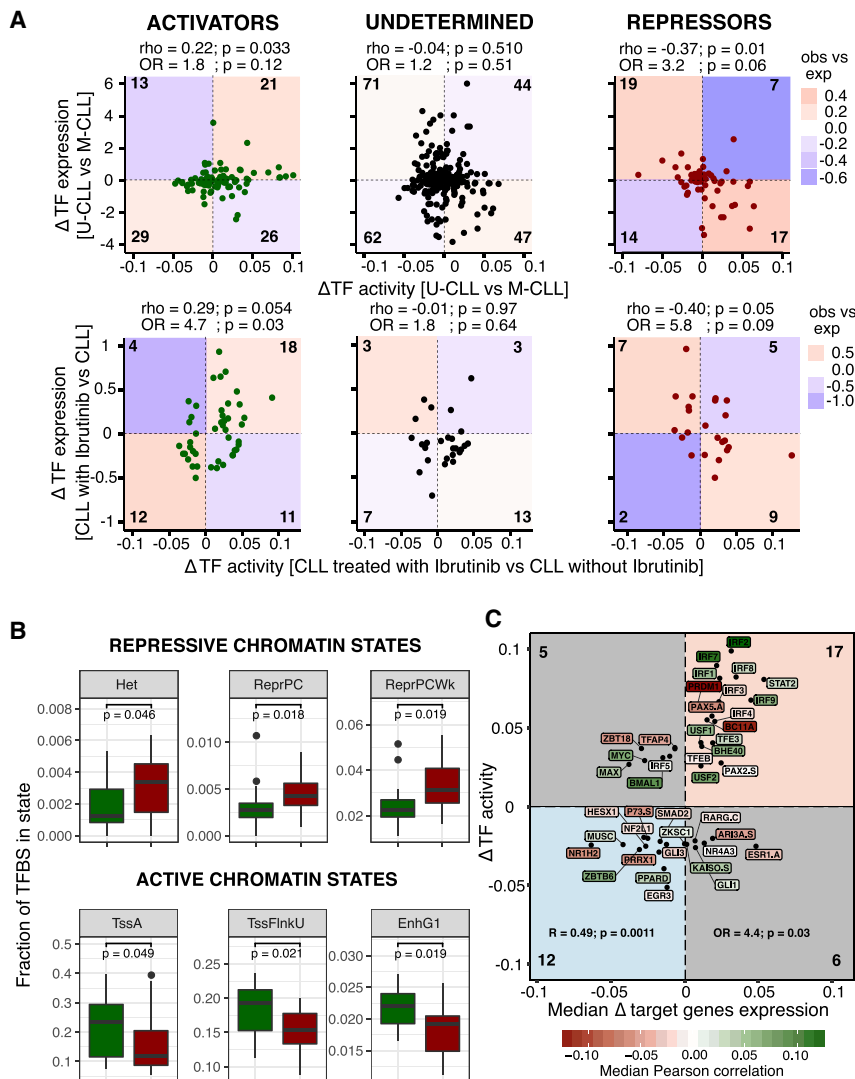


Figure 4. Experimental Validation for the Activator/Repressor Classification in Case Study I.

(A) Correlation of differential TF activity and differential gene expression (log2 fold changes obtained with DESeq) for predicted activators (left), undetermined TFs (middle), and predicted repressors (right) for the comparison of U-CLL and M-CLL (top) and for CLL samples treated with ibrutinib versus control treatment with DMSO (bottom). TF classifications were obtained from data of case study I. Ibrutinib treatment was performed on an independent patient cohort ($n = 4$). Spearman's ρ and p value, as well as the odds ratio (OR) and p value of Fisher's exact test, are reported in the figure. The number (n) of TFs is indicated for each quadrant. For the bottom row, only TFs that were significant in diffTF in case study I ($FDR < 10\%$) are shown. Color shadings indicate the observed versus the expected ratio for each quadrant (blue, less than expected; red, more than expected).

(B) Fraction of TFBSs overlapping specific chromatin states (Kundaje et al., 2015) are shown for putative activators (green) and repressors (red). Only chromatin states with significant differences between activators and repressors are displayed (Wilcoxon test, $p < 0.05$).

(C) Correlation of differential TF activity (diffTF U-CLL versus M-CLL) and the differential expression of target genes (median log2 fold change U-CLL versus M-CLL) are shown. Color of TF labels represents mode-of-action class (activator, green; repressor, red) on a continuous scale based on the correlation strength (odds ratio [OR] and p value are given for Fisher's Exact test; R and corresponding p value are given for Pearson's correlation).

The data presented in this figure are based on 27 and 25 biological replicates for U-CLL and M-CLL, respectively.

regardless of whether the TF was classified as activator or repressor (green or red, respectively, in Figure 4C), thus demonstrating the correspondence between chromatin-based prediction of the mode of action and its actual effect on the expression of the target genes.

In summary, we show that diffTF is able to classify TFs by their mode of action and that the activity of most TFs follows the predicted direction upon perturbation of their expression level. TFs that did not follow the expected trend may represent interesting biology, e.g., posttranslational regulation, such that a change in expression would not reflect a change in the abundance of its active form. Future studies may use this information for investigating different ways of regulating the activity of a TF.

Case Study II: Applying diffTF to a Small-Scale Multiomics Dataset

A common study design is to assess the differences between two distinct conditions with few biological replicates per condition. To assess the applicability of diffTF for such data, we

applied it to the well-studied mouse hematopoietic system. We used eight ATAC-seq and RNA-seq profiles of multipotent progenitor (MPP) cells ($\text{Lin}^- \text{cKit}^+ \text{Sca1}^+$; $\text{CD150}^+ \text{CD48}^+$), an early hematopoietic progenitor population capable of supporting multilineage blood production (Sun et al., 2014), and compared them with the more differentiated and myeloid-restricted GMPs ($\text{Lin}^- \text{cKit}^+ \text{Sca1}^-$; CD16/32^+). The data, which are also published as part of our previous study (Rasmussen et al., 2019), were processed using our in-house ATAC-seq pipeline (see STAR Methods). diffTF (using the analytical procedure; see STAR Methods) identified the well-known class of master regulators of myelopoiesis, the CEBP family ($\text{C/ebp}\alpha$, $\text{C/ebp}\beta$, $\text{C/ebp}\delta$, $\text{C/ebp}\epsilon$, and $\text{C/ebp}\gamma$) as the most differentially active TFs in GMPs versus MPPs (Figure 5A; Figures S6A and S6B). In addition, we observed higher activity of the MYC/MYB factors, known to be exclusively active in GMPs (Baker et al., 2014), and NFIL3, which is involved in the generation of natural killer cells (Gascoyne et al., 2009). Conversely, MPPs showed higher activity for IRF/STAT, ZEB1, and ITF2 (part of Wnt signaling),

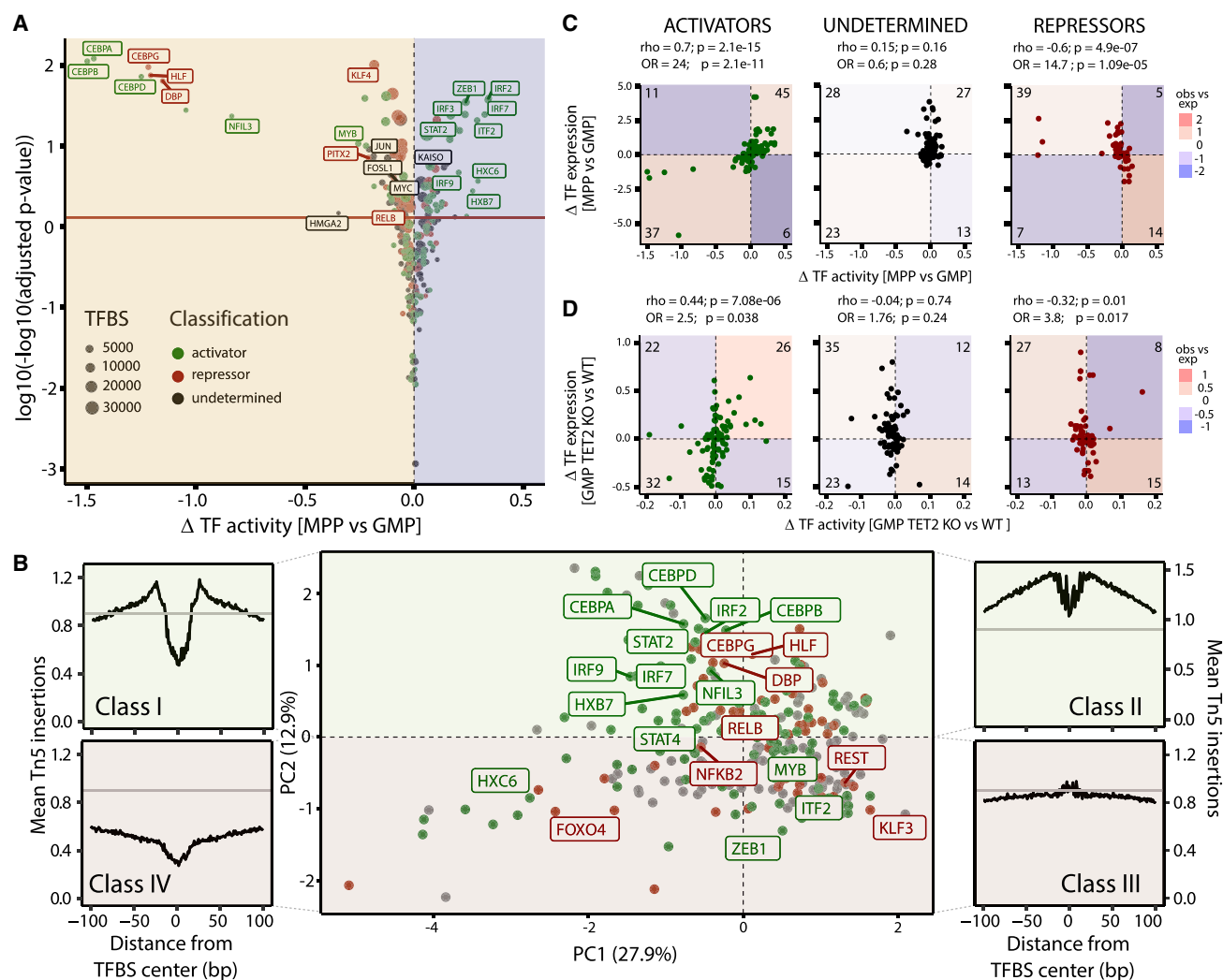


Figure 5. Results of diffTF in Basic and Classification Mode Applied to Case Study II

TFs are colored according to their predicted classification (red, repressors; gray/black, undetermined; green, activators). Only expressed TFs are shown ($n = 268$). (A) Volcano plot of differential TF activity between MPPs ($n = 4$ biological replicates) and GMPs ($n = 4$ biological replicates). Only the most significant TFs that are relevant for hematopoietic stem cell differentiation are labeled (see Table S1 for full list). P values are obtained through diffTF using the analytical approach and adjusted by the Benjamini-Hochberg procedure (y-axis).

(B) PC1 versus PC2 of a PCA on footprint data in analogy to Figure 3D. Classes are labeled based on Figure 3D.

(C and D) Correlation of differential TF activity and differential expression (\log_2 fold changes obtained from DESeq) are shown for activators (left), undetermined TFs (middle), and repressors (right) for the comparison of GMP versus MPP (C) and GMP TET2 knockout versus wild type, which represents independent data (D). Color shadings in (C) and (D) indicate the observed versus expected \log_2 ratio for each quadrant (blue, less than expected; red, more than expected). TF classifications were obtained from the wild-type GMP and MPP samples. Spearman's rho and p value, as well as the odds ratio (OR) and p values of Fisher's exact test, are provided. The number of TFs is indicated in each quadrant.

as well as TFs from the Homeodomain (HXB7 and HXA10) and Forkhead (FOXO3) families, all of which are associated with self-renewal of HSCs (Sands et al., 2013).

Applying diffTF in classification mode classified 158 TFs (of 268 expressed TFs) as putative activators or repressors. A PCA of the TF footprints revealed, similar to case study I, that PC1 separates high from low motif accessibility and PC2 separates high from low flank accessibility, thus suggesting this as a general pattern (Figure 5B). The most differentially active TFs between MPP and GMP were classified as activators (CEBPs,

NFIL3, and IRFs) or had overall mixed evidence (i.e., activator footprint but inconsistent correlation direction of expression and activity, such as DBP and HLF). The magnitude of differential activity of the most differential putative repressor, JUN, was far below that of the differential activators, indicating that the differentiation process from MPP to GMP is mainly driven by transcriptional activators. We again validated the classifications by comparing differential activity and differential expression (MPP versus GMP) and found overall more positive correlations for putative activators ($\rho = 0.69$, $p = 2.1 \times 10^{-15}$) and more negative

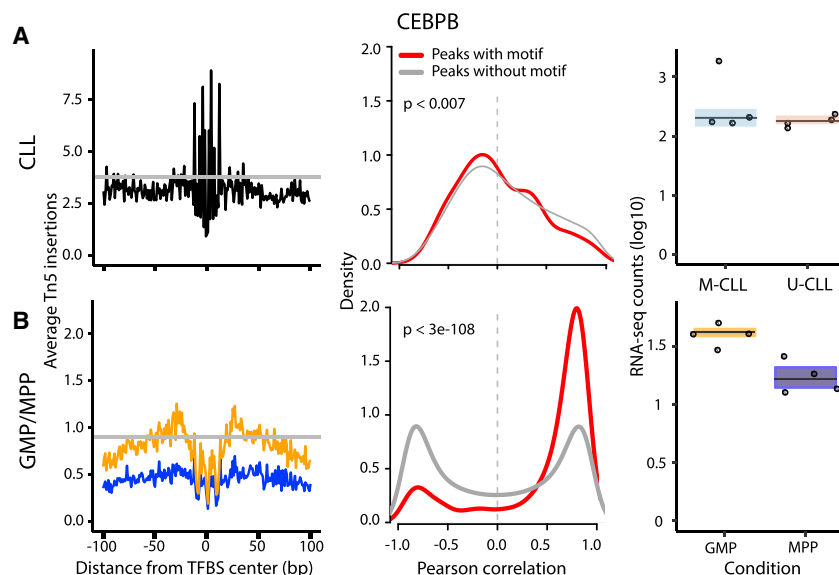


Figure 6. Comparison of diffTF Results of Case Study I and II

Footprints (left), correlation distributions (middle), and expression values (log10-transformed RNA-seq counts, right) are shown for CEBPB in CLL data from case study I, where it was classified as putative repressor (A), and in the GMP (blue) and MPP (yellow) data from case study II, where it was classified as activator (B). The p value for both (A) and (B) are given for the Wilcoxon Rank sum test.

qualitatively (chromVAR, BagFootT, and HOMER; Table S4) and quantitatively (HOMER and chromVAR). Overall, these comparisons revealed that although the results between tools are comparable (see STAR Methods), diffTF seems more sensitive than, e.g., HOMER, and it provides a more flexible and tailored analysis framework than chromVAR (which was developed for single-cell ATAC-seq data)

correlations for putative repressors ($\rho = -0.58$, $p = 4.9 \times 10^{-7}$; Figure 5C). The change in TF activity based on chromatin accessibility was again correlated with the change in target gene expression, regardless of whether the TF was classified as a repressor or an activator (Figure S6D).

For the independent validation of the classification, we obtained data from our earlier work (Rasmussen et al., 2019), which measured the effect of TET2 knockout (KO) in GMP cells. We calculated the correlations between differential expression and differential TF activity for TET2 wild type versus KO while stratifying the TFs based on the classes from case study II (GMP and MPP in wild type). We found that 100 of 158 TFs (63%) behaved in the direction as predicted by their class (Figure 5D).

Overall, these results show that diffTF is able to identify the known TFs that drive the differentiation from MPP to GMP, thus demonstrating its power to detect signals for study designs with a small number of replicates. Furthermore, we show that our classification scheme correctly predicts the directionality of change in TF activity upon genetic perturbation in most TFs.

Having performed the activator/repressor classification in two settings within the hematopoietic differentiation system, we quantified how many TFs changed their mode of action between the two case studies. In total, we found 16 TFs were classified as potential activators in case study I and repressors in case study II, or vice versa (Table S1). Among these, CEBPB showed the most divergent footprints, which were in the same direction as the diffTF classifications in each case study: It is classified as a repressor and has a repressor-like footprint in CLL (Figure 6A) and is classified as an activator with an activator-like footprint in GMP/MPP (Figure 6B). Because CEBPB is one of the TFs driving differentiation in the hematopoietic system, we speculate that its role as a repressor in CLL may contribute to the malignancy of the disease.

Comparison with Similar Tools

Finally, we compared diffTF to the few tools with a similar focus (Baek et al., 2017; Heinz et al., 2010; Schep et al., 2017), both

because of the extensive choice of parameters, diagnostic plots, TFBS-specific results, visualizations, and pipeline adjustability. Its most outstanding feature is its ability to directly integrate RNA-seq with ATAC-seq data for classifying TFs into putative activators and repressors.

DISCUSSION

We presented a genome-wide method for quantifying differential TF activity for a large set of TFs simultaneously and for classifying them into their molecular mode of action as putative transcriptional activators or repressors. The method, which is available for download at <https://git.embl.de/grp-zaugg/diffTF>, along with comprehensive documentation and example data, has been applied in the context of two case studies that demonstrate its power to uncover biologically meaningful changes in TF activity in (1) a heterogeneous patient cohort of two subtypes of a disease (U-CLL versus M-CLL) and (2) in a small-scale cell-type comparison with few but homogeneous biological replicates (GMP versus MPP in isogenic mice). We extensively tested and demonstrated the technical robustness of diffTF and showed that it is able to overcome the inherent noisiness of TF binding site predictions by aggregating data across all putative binding sites.

Calculating differential TF activity based on aggregating signals across the genome has been proposed before based on the expression of putative target genes of a certain TF (Boorsma et al., 2008; Bussemaker et al., 2001). However, using chromatin instead of gene expression to calculate TF activity has several advantages: First, chromatin is a simpler trait than gene expression (which is the result of transcription and degradation) and thus more sensitive to detect small differences. Second, there are many more peaks than genes, thus allowing better statistics and signal-to-noise ratio. Third, the effect on chromatin is locally defined, whereas the elements regulating gene expression (e.g., enhancers) are often not well defined (i.e., enhancer-gene linking is an active area of research).

In classification mode, diffTF can classify TFs into putative activators and repressors based on the correlation of their expression level (RNA-seq) and the activity of their putative binding sites (ATAC-seq). This information is relevant when interpreting the action of TFs, because it is important to know whether upregulation of a TF has a positive or negative effect on chromatin (and therefore transcriptional) activity. Our experimental validations demonstrate that diffTF can correctly predict the directionality of the effect of transcriptional upregulation of TFs on their chromatin activity upon chemical stimulation (in case study I) or genetic perturbation (in case study II) for most TFs. Nevertheless, we note that diffTF is prone to mis-classifying TFs that (1) act bifunctionally as activators and repressors, (2) are heavily regulated posttranslationally, or (3) show little variation in RNA expression across the samples. Some of these mis-classifications may represent interesting subjects for future investigations. Furthermore, if two TFs have similar motifs, which makes it difficult to distinguish them, diffTF may have difficulties in classifying them correctly. Thus, for distinguishing the functional roles of TFs from the same motif-family, further biochemical experiments are needed. Despite these potential pitfalls, diffTF will be highly relevant for generating hypotheses about the molecular mechanism of TFs on a global level.

Although similar methods have been proposed for analyzing ATAC-seq data (Baek et al., 2017; Schep et al., 2017), our method has several advantages when dealing with bulk ATAC-seq data and can be used for histone mark ChIP-seq data: (1) Unlike other methods that calculate the background theoretically based on the genome-wide read depth, diffTF is insensitive to sequence and locus-dependent biases, because we calculate a fold change between conditions for each region, thus normalizing for local read depth biases. This is particularly advantageous for detecting small changes such as those between two heterogeneous cohorts in patient-control studies. (2) diffTF allows integration with matching RNA-seq data to classify TFs into activators and repressors in a fully data-driven approach within the same analysis framework. Such classifications are crucial when interpreting the effects of up/downregulation of a particular factor. (3) diffTF provides the fold change value for each TFBS, which allows easy retrospective follow-up analysis, e.g., identifying the most differential regions regulated by a specific set of TFs. (4) Our method provides the possibility to analyze time course data additively by calculating the overall change of slope for each TF (see STAR Methods). Because of its flexibility, diffTF is computationally expensive, and we provide detailed instructions on memory and time requirements in the documentation.

Overall, we presented a multiomics data integration strategy of ATAC-seq and RNA-seq data that calculates differential TF activity across conditions and classifies TFs based on their molecular mode of action. The main goal of diffTF is to aid in formulating testable hypotheses and ultimately improve the understanding of regulatory mechanisms that are driving differences in cell state on a system-wide scale.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **LEAD CONTACT AND MATERIALS AVAILABILITY**
- **EXPERIMENTAL MODEL AND SUBJECT DETAILS**
- **METHOD DETAILS**
 - Summary of publicly available data used in analyses
 - Tools used
 - ATAC-Seq processing
 - Description of the diffTF analysis workflow
 - Validation of the analytical approach
 - Guidance on the number of permutations and whether to use the analytical or permutation-based approach
 - Classification of TFs into activator and repressors
 - CLL ATAC-Seq processing
 - Running diffTF on the CLL data
 - Comparison of diffTF results based on predicted versus ChIP-Seq validated TFBS
 - Assessing impact of TF motif scanning parameters on diffTF results
 - diffTF robustness analysis with respect to internal parameters for the CLL data
 - Assessing dependence of diffTF results on differential signal strength for CLL data
 - GO terms associated with CLL progression
 - Assessment of the power of diffTF with respect to sample size and sequencing depth for CLL data
 - RNA-seq processing for CLL data and classification into activators and repressors
 - Correlation of TF activity with target genes expression for CLL data
 - Footprinting analysis for CLL data
 - Chromatin state enrichment in CLL data
 - CLL data treated with Ibrutinib
 - ATAC-seq libraries generation of the CLL treated with Ibrutinib
 - RNA-seq library generation for the CLL dataset treated with Ibrutinib
 - Comparing diffTF results with HOMER for CLL data
 - Comparison with similar tools
 - HSC mouse data source (FACS sorting step)
 - HSC ATAC-Seq libraries generation
 - GMP-MPP ATAC-Seq processing
 - GMP-MPP diffTF analysis
 - GMP-MPP footprinting analysis
- **QUANTIFICATION AND STATISTICAL ANALYSIS**
- **DATA AND CODE AVAILABILITY**

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.celrep.2019.10.106>.

ACKNOWLEDGMENTS

We thank Bernd Klaus for help with the statistical part and EMBL for funding. A.R.-P. is the recipient of a postdoctoral fellowship granted by Fundación Ramón Areces. G.P. was supported by the Otto Bayer Scholarship from Bayer Foundations. We thank EMBL GeneCore for help with the sequencing. K.D.R. was supported by a postdoctoral fellowship from the Danish Medical Research Council (FSS 1333-00120B). The work in the Helin

laboratory was supported by grants to K.H. from the Independent Research Fund Denmark (8020-00044B), from the European Research Council (294666_DNAME), through a center grant from the Novo Nordisk Foundation (NNF17CC0027852), and through a Memorial Sloan Kettering Cancer Center support grant (NIH P30 CA008748).

AUTHOR CONTRIBUTIONS

J.B.Z. and I.B. conceived the study; C.A. and I.B. developed the computational framework and performed the analyses; J.B.Z. supervised the tool development and analysis; A.R.-P. and G.P. contributed to the development and analysis; K.D.R., H.G., and P.-M.B. performed experiments; K.H., S.D., and W.H. supervised the experiments; J.B.Z. supervised the study; and C.A., I.B., and J.B.Z. wrote the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: December 1, 2018

Revised: September 20, 2019

Accepted: October 28, 2019

Published: December 3, 2019

REFERENCES

- Ambrosini, G., Groux, R., and Bucher, P. (2018). PWMScan: a fast tool for scanning entire genomes with a position-specific weight matrix. *Bioinformatics* 34, 2483–2484.
- Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>.
- Arvaniti, E., Ntoufa, S., Papakonstantinou, N., Touloumenidou, T., Laoutaris, N., Anagnostopoulos, A., Lamnissou, K., Caligaris-Cappio, F., Stamatoopoulos, K., Ghia, P., et al. (2011). Toll-like receptor signaling pathway in chronic lymphocytic leukemia: distinct gene expression profiles of potential pathogenic significance in specific subsets of patients. *Haematologica* 96, 1644–1652.
- Baek, S., Goldstein, I., and Hager, G.L. (2017). Bivariate genomic footprinting detects changes in transcription factor activity. *Cell Rep.* 19, 1710–1722.
- Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W., and Noble, W.S. (2009). MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* 37, W202–W208.
- Baker, S.J., Ma'ayan, A., Lieu, Y.K., John, P., Reddy, M.V.R., Chen, E.Y., Duan, Q., Snoeck, H.-W., and Reddy, E.P. (2014). B-myb is an essential regulator of hematopoietic stem cell and myeloid progenitor cell development. *Proc. Natl. Acad. Sci. USA* 111, 3122–3127.
- Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* 57, 289–300.
- Benjamini, Y., and Speed, T.P. (2012). Summarizing and correcting the GC content bias in high-throughput sequencing. *Nucleic Acids Res.* 40, e72.
- Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120.
- Bonhoure, N., Bounova, G., Bernasconi, D., Praz, V., Lammers, F., Canella, D., Willis, I.M., Herr, W., Hernandez, N., and Delorenzi, M.; CycliX Consortium (2014). Quantifying ChIP-seq data: a spiking method providing an internal reference for sample-to-sample normalization. *Genome Res.* 24, 1157–1168.
- Boorsma, A., Lu, X.-J., Zakrzewska, A., Klis, F.M., and Bussemaker, H.J. (2008). Inferring condition-specific modulation of transcription factor activity in yeast through regulon-based analysis of genomewide expression. *PLoS ONE* 3, e3112.
- Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* 10, 1213–1218.

- Bussemaker, H.J., Li, H., and Siggia, E.D. (2001). Regulatory element detection using correlation with expression. *Nat. Genet.* 27, 167–171.
- Castro-Mondragon, J.A., Jaeger, S., Thieffry, D., Thomas-Chollier, M., and van Helden, J. (2017). RSAT matrix-clustering: dynamic exploration and redundancy reduction of transcription factor binding motif collections. *Nucleic Acids Res.* 45, e119.
- Chevrier, S., Emslie, D., Shi, W., Kratina, T., Wellard, C., Karnowski, A., Erikci, E., Smyth, G.K., Chowdhury, K., Tarlinton, D., and Corcoran, L.M. (2014). The BTB-ZF transcription factor Zbtb20 is driven by Irf4 to promote plasma cell differentiation and longevity. *J. Exp. Med.* 211, 827–840.
- Chiorazzi, N., and Ferrarini, M. (2011). Cellular origin(s) of chronic lymphocytic leukemia: cautionary notes and additional considerations and possibilities. *Blood* 117, 1781–1791.
- Coscia, M., Pantaleoni, F., Riganti, C., Vitale, C., Rigoni, M., Peola, S., Castella, B., Foglietta, M., Griggio, V., Drandi, D., et al. (2011). IGHV unmutated CLL B cells are more prone to spontaneous apoptosis and subject to environmental prosurvival signals than mutated CLL B cells. *Leukemia* 25, 828–837.
- D'Annibale, S., Kim, J., Magliozzi, R., Low, T.Y., Mohammed, S., Heck, A.J.R., and Guardavaccaro, D. (2014). Proteasome-dependent degradation of transcription factor activating enhancer-binding protein 4 (TFAP4) controls mitotic division. *J. Biol. Chem.* 289, 7730–7737.
- Dietrich, S., Oleś, M., Lu, J., Sellner, L., Anders, S., Velten, B., Wu, B., Hülle, J., da Silva Liberio, M., Walther, T., et al. (2018). Drug-perturbation-based stratification of blood cancer. *J. Clin. Invest.* 128, 427–445.
- Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21.
- El-Athman, R., and Relógio, A. (2018). Escaping circadian regulation: an emerging hallmark of cancer? *Cell Syst.* 6, 266–267.
- Ernst, J., and Kellis, M. (2012). ChromHMM: automating chromatin-state discovery and characterization. *Nat. Methods* 9, 215–216.
- Ewels, P., Magnusson, M., Lundin, S., and Käller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* 32, 3047–3048.
- Frankish, A., Diekhans, M., Ferreira, A.-M., Johnson, R., Jungreis, I., Loveland, J., Mudge, J.M., Sisu, C., Wright, J., Armstrong, J., et al. (2019). GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res.* 47 (D1), D766–D773.
- Furman, R.R., Sharman, J.P., Coutre, S.E., Cheson, B.D., Pagel, J.M., Hillmen, P., Barrientos, J.C., Zelenetz, A.D., Kipps, T.J., Flinn, I., et al. (2014). Idelalisib and rituximab in relapsed chronic lymphocytic leukemia. *N. Engl. J. Med.* 370, 997–1007.
- Gascoyne, D.M., Long, E., Veiga-Fernandes, H., de Boer, J., Williams, O., Seddon, B., Coles, M., Kioussis, D., and Brady, H.J.M. (2009). The basic leucine zipper transcription factor E4BP4 is essential for natural killer cell development. *Nat. Immunol.* 10, 1118–1124.
- Griffon, A., Barbier, Q., Dalino, J., van Helden, J., Spicuglia, S., and Ballester, B. (2015). Integrative analysis of public ChIP-seq experiments reveals a complex multi-cell regulatory landscape. *Nucleic Acids Res.* 43, e27.
- Grubert, F., Zugg, J.B., Kasowski, M., Ursu, O., Spacek, D.V., Martin, A.R., Greenside, P., Srivas, R., Phanstiel, D.H., Pekowska, A., et al. (2015). Genetic control of chromatin states in humans involves local and distal chromosomal interactions. *Cell* 162, 1051–1065.
- Grüning, B., Dale, R., Sjödin, A., Chapman, B.A., Rowe, J., Tomkins-Tinch, C.H., Valieris, R., and Köster, J.; Bioconda Team (2018). Bioconda: sustainable and comprehensive software distribution for the life sciences. *Nat. Methods* 15, 475–476.
- Han, H., Shim, H., Shin, D., Shim, J.E., Ko, Y., Shin, J., Kim, H., Cho, A., Kim, E., Lee, T., et al. (2015). TRRUST: a reference database of human transcriptional regulatory interactions. *Sci. Rep.* 5, 11432.
- Han, H., Cho, J.-W., Lee, S., Yun, A., Kim, H., Bae, D., Yang, S., Kim, C.Y., Lee, M., Kim, E., et al. (2018). TRRUST v2: an expanded reference database of

- p human and mouse transcriptional regulatory interactions.
- Nucleic Acids Res.*
- 46 (D1), D380–D386.
- Harrow, J., Frankish, A., Gonzalez, J.M., Tapanari, E., Diekhans, M., Kokocinski, F., Aken, B.L., Barrell, D., Zadissa, A., Searle, S., et al. (2012). GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res.* 22, 1760–1774.
- Havelange, V., Pekarsky, Y., Nakamura, T., Palamarchuk, A., Alder, H., Rasenti, L., Kipps, T., and Croce, C.M. (2011). IRF4 mutations in chronic lymphocytic leukemia. *Blood* 118, 2827–2829.
- Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* 38, 576–589.
- Huber, W., Carey, V.J., Gentleman, R., Anders, S., Carlson, M., Carvalho, B.S., Bravo, H.C., Davis, S., Gatto, L., Girke, T., et al. (2015). Orchestrating high-throughput genomic analysis with Bioconductor. *Nat. Methods* 12, 115–121.
- Kasim, V., Xie, Y.-D., Wang, H.-M., Huang, C., Yan, X.-S., Nian, W.-Q., Zheng, X.-D., Miyagishi, M., and Wu, S.-R. (2017). Transcription factor Yin Yang 2 is a novel regulator of the p53/p21 axis. *Oncotarget* 8, 54694–54707.
- Kauffman, E.C., Ricketts, C.J., Rais-Bahrami, S., Yang, Y., Merino, M.J., Bottaro, D.P., Srinivasan, R., and Linehan, W.M. (2014). Molecular genetics and cellular features of TFE3 and TFEB fusion kidney cancers. *Nat. Rev. Urol.* 11, 465–475.
- Kern, D., Regl, G., Hofbauer, S.W., Altenhofer, P., Achatz, G., Dlugosz, A., Schnidar, H., Greil, R., Hartmann, T.N., and Aberger, F. (2015). Hedgehog/GLI and PI3K signaling in the initiation and maintenance of chronic lymphocytic leukemia. *Oncogene* 34, 5341–5351.
- Kikushige, Y., Ishikawa, F., Miyamoto, T., Shima, T., Urata, S., Yoshimoto, G., Mori, Y., Iino, T., Yamauchi, T., Eto, T., et al. (2011). Self-renewing hematopoietic stem cell is the primary target in pathogenesis of human chronic lymphocytic leukemia. *Cancer Cell* 20, 246–259.
- Kim, R., Emi, M., and Tanabe, K. (2007). Cancer immunoediting from immune surveillance to immune escape. *Immunology* 121, 1–14.
- Komatsu, M., Kurokawa, H., Waguri, S., Taguchi, K., Kobayashi, A., Ichimura, Y., Sou, Y.-S., Ueno, I., Sakamoto, A., Tong, K.I., et al. (2010). The selective autophagy substrate p62 activates the stress responsive transcription factor Nrf2 through inactivation of Keap1. *Nat. Cell Biol.* 12, 213–223.
- Köster, J., and Rahmann, S. (2012). Snakemake—a scalable bioinformatics workflow engine. *Bioinformatics* 28, 2520–2522.
- Kreslavsky, T., Vilagos, B., Tagoh, H., Poliakov, D.K., Schwickert, T.A., Wöhner, M., Jaritz, M., Weiss, S., Taneja, R., Rossner, M.J., and Busslinger, M. (2017). Essential role for the transcription factor Bhlhe41 in regulating the development, self-renewal and BCR repertoire of B-1a cells. *Nat. Immunol.* 18, 442–455.
- Kulakovskiy, I.V., Medvedeva, Y.A., Schaefer, U., Kasianov, A.S., Vorontsov, I.E., Bajic, V.B., and Makeev, V.J. (2013). HOCOMOCO: a comprehensive collection of human transcription factor binding sites models. *Nucleic Acids Res.* 41, D195–D202.
- Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., Ziller, M.J., et al.; Roadmap Epigenomics Consortium (2015). Integrative analysis of 111 reference human epigenomes. *Nature* 518, 317–330.
- Landau, D.A., Tausch, E., Taylor-Weiner, A.N., Stewart, C., Reiter, J.G., Bahlo, J., Kluth, S., Bozic, I., Lawrence, M., Böttcher, S., et al. (2015). Mutations driving CLL and their evolution in progression and relapse. *Nature* 526, 525–530.
- Landt, S.G., Marinov, G.K., Kundaje, A., Kheradpour, P., Pauli, F., Batzoglou, S., Bernstein, B.E., Bickel, P., Brown, J.B., Cayting, P., et al. (2012). ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res.* 22, 1813–1831.
- Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359.
- Lara-Astiaso, D., Weiner, A., Lorenzo-Vivas, E., Zaretzky, I., Jaitin, D.A., David, E., Keren-Shaul, H., Mildner, A., Winter, D., Jung, S., et al. (2014). Immunogenetics. Chromatin state dynamics during blood formation. *Science* 345, 943–949.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079.
- Li, Y.-J., Sun, L., Shi, Y., Wang, G., Wang, X., Dunn, S.E., Iorio, C., Screaton, R.A., and Spaner, D.E. (2017). PPAR-delta promotes survival of chronic lymphocytic leukemia cells in energetically unfavorable conditions. *Leukemia* 31, 1905–1914.
- Liao, Y., Smyth, G.K., and Shi, W. (2013). The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res.* 41, e108.
- Liu, Z., Lee, J., Krummey, S., Lu, W., Cai, H., and Lenardo, M.J. (2011). The kinase LRRK2 is a regulator of the transcription factor NFAT that modulates the severity of inflammatory bowel disease. *Nat. Immunol.* 12, 1063–1070.
- Liu, L., Jin, G., and Zhou, X. (2015). Modeling the relationship of epigenetic modifications to transcription factor binding. *Nucleic Acids Res.* 43, 3873–3885.
- Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550.
- Lun, A.T.L., and Smyth, G.K. (2016). csaw: a Bioconductor package for differential binding analysis of ChIP-seq data using sliding windows. *Nucleic Acids Res.* 44, e45.
- Mathelier, A., Fornes, O., Arenillas, D.J., Chen, C.-Y., Denay, G., Lee, J., Shi, W., Shyr, C., Tan, G., Worsley-Hunt, R., et al. (2016). JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* 44 (D1), D110–D115.
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernysky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M.A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303.
- Mezger, A., Klemm, S., Mann, I., Brower, K., Mir, A., Bostick, M., Farmer, A., Fordyce, P., Linnarsson, S., and Greenleaf, W. (2018). High-throughput chromatin accessibility profiling at single-cell resolution. *Nat. Commun.* 9, 3647.
- Minami, Y., Oishi, I., Endo, M., and Nishita, M. (2010). Ror-family receptor tyrosine kinases in noncanonical Wnt signaling: their implications in developmental morphogenesis and human diseases. *Dev. Dyn.* 239, 1–15.
- Oakes, C.C., Seifert, M., Assenov, Y., Gu, L., Przekopowicz, M., Ruppert, A.S., Wang, Q., Imbusch, C.D., Serva, A., Koser, S.D., et al. (2016). DNA methylation dynamics during B cell maturation underlie a continuum of disease phenotypes in chronic lymphocytic leukemia. *Nat. Genet.* 48, 253–264.
- Piper, J., Assi, S.A., Cauchy, P., Ladrone, C., Cockerill, P.N., Bonifer, C., and Ott, S. (2015). Wellington-bootstrap: differential DNase-seq footprinting identifies cell-type determining transcription factors. *BMC Genomics* 16, 1000.
- Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842.
- Ramírez, F., Dündar, F., Diehl, S., Grüning, B.A., and Manke, T. (2014). deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res.* 42, W187–W191.
- Rasmussen, K.D., Berest, I., Keßler, S., Nishimura, K., Simón-Carrasco, L., Vassiliou, G.S., Pedersen, M.T., Christensen, J., Zaugg, J.B., and Helin, K. (2019). TET2 binding to enhancers facilitates transcription factor recruitment in hematopoietic cells. *Genome Res.* 29, 564–575.
- Rendeiro, A.F., Schmidl, C., Strefford, J.C., Walewska, R., Davis, Z., Farlik, M., Oscier, D., and Bock, C. (2016). Chromatin accessibility maps of chronic lymphocytic leukaemia identify subtype-specific epigenome signatures and transcription regulatory networks. *Nat. Commun.* 7, 11938.
- Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 43, e47.

- Ross-Innes, C.S., Stark, R., Teschendorff, A.E., Holmes, K.A., Ali, H.R., Dunning, M.J., Brown, G.D., Gojis, O., Ellis, I.O., Green, A.R., et al. (2012). Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature* **481**, 389–393.
- Saito, T., and Rehmsmeier, M. (2017). Precrec: fast and accurate precision-recall and ROC curve calculations in R. *Bioinformatics* **33**, 145–147.
- Sands, W.A., Copland, M., and Wheadon, H. (2013). Targeting self-renewal pathways in myeloid malignancies. *Cell Commun. Signal.* **11**, 33.
- Schep, A.N., Wu, B., Buenrostro, J.D., and Greenleaf, W.J. (2017). chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nat. Methods* **14**, 975–978.
- Slager, S.L., Caporaso, N.E., de Sanjose, S., and Goldin, L.R. (2013). Genetic susceptibility to chronic lymphocytic leukemia. *Semin. Hematol.* **50**, 296–302.
- Sun, J., Ramos, A., Chapman, B., Johnnidis, J.B., Le, L., Ho, Y.-J., Klein, A., Hofmann, O., and Camargo, F.D. (2014). Clonal dynamics of native haematopoiesis. *Nature* **514**, 322–327.
- Teng, M.W.L., Swann, J.B., Koebel, C.M., Schreiber, R.D., and Smyth, M.J. (2008). Immune-mediated dormancy: an equilibrium with cancer. *J. Leukoc. Biol.* **84**, 988–993.
- Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B., Lee, T.I., and Young, R.A. (2013). Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307–319.
- Worsley Hunt, R., Mathelier, A., Del Peso, L., and Wasserman, W.W. (2014). Improving analysis of transcription factor binding sites within ChIP-Seq data based on topological motif enrichment. *BMC Genomics* **15**, 472.
- Yeomans, A., Thirdborough, S.M., Valle-Argos, B., Linley, A., Krysov, S., Hidalgo, M.S., Leonard, E., Ishfaq, M., Wagner, S.D., Willis, A.E., et al. (2016). Engagement of the B-cell receptor of chronic lymphocytic leukemia cells drives global and MYC-specific mRNA translation. *Blood* **127**, 449–457.
- Yu, G., Wang, L.-G., and He, Q.-Y. (2015). ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* **31**, 2382–2383.
- Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137.
- Zhou, Y., Li, Y.-S., Bandi, S.R., Tang, L., Shinton, S.A., Hayakawa, K., and Hardy, R.R. (2015). Lin28b promotes fetal B lymphopoiesis through the transcription factor Arid3a. *J. Exp. Med.* **212**, 569–580.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological Samples		
Peripheral blood samples	Heidelberg University Hospital	N/A
Chemicals, Peptides, and Recombinant Proteins		
RPMI	GIBCO	Cat.No. 21875-034
Glutamine	GIBCO	Cat.No. 25030-24
Pen/Strep	GIBCO	Cat.No. 15140-122
Ibrutinib	Selleckchem	Cat.No. S2680
DMSO	SERVA	Cat.No. 20385
NEBNext High-Fidelity 2x PCR Master Mix	New England Biolabs	NEB #M0544
Critical Commercial Assays		
Nextera index kit	Illumina	Cat.No. 20027213
DNA High Sensitivity kit	Agilent Technologies	Cat.No. 5067-4626
miRNeasy Mini Kit	QIAGEN	Cat.No. 217004
RNA Nano 6000 Assay Kit	Agilent Technologies	Cat.No. 5067-1511
TruSeq RNA Sample Preparation v2 Kit	Illumina	RS-122-2001
Deposited Data		
Raw human CLL ATAC-seq and RNA-seq data	Rendeiro et al., 2016	EGA: EGAS00001001821
TET2 knockout mouse ATAC-seq and RNA-seq data	Rasmussen et al., 2019	GEO: GSE115972
Human reference genome NCBI build 37, GRCh37	Genome Reference Consortium	http://www.ncbi.nlm.nih.gov/projects/genome/assembly/grc/human/
Mouse reference genome NCBI build 38, GRCm38	Genome Reference Consortium	http://www.ncbi.nlm.nih.gov/projects/genome/assembly/grc/human/
Software and Algorithms		
diffTF	this paper	https://diffTF.readthedocs.io/en/latest/
Snakemake 5.0	Köster and Rahmann, 2012	https://snakemake.readthedocs.io/ ; RRID:SCR_003475
Bioconda project	Grüning et al., 2018	https://bioconda.github.io/
Bioconductor	Huber et al., 2015	https://www.bioconductor.org/ ; RRID:SCR_006442
Conda package manager	Continuum Analytics, Inc.	https://docs.conda.io/en/latest/
bedtools 2.26.0	Quinlan and Hall, 2010	https://bedtools.readthedocs.io/en/latest/ ; RRID:SCR_006646
samtools 1.4	Li et al., 2009	http://www.htslib.org/ ; RRID:SCR_002105
Bowtie2 2.3.0	Langmead and Salzberg, 2012	http://bowtie-bio.sourceforge.net/bowtie2
deepTools 2.5.0	Ramírez et al., 2014	https://deeptools.readthedocs.io/en/develop/ ; RRID:SCR_016366
FastQC 0.11.5	Andrews, 2010	https://github.com/s-andrews/FastQC ; RRID:SCR_014583
GATK 3.7	McKenna et al., 2010	https://software.broadinstitute.org/gatk/ ; RRID:SCR_001876
pyDNase 0.2.4	Piper et al., 2015	https://pythonhosted.org/pyDNase/ ; RRID:SCR_005406
STAR 2.4.1	Dobin et al., 2013	https://github.com/alexdobin/STAR ; RRID:SCR_015899
MACS2 2.1.1	Zhang et al., 2008	https://github.com/taoliu/MACS ; RRID:SCR_013291
multiqc 0.9	Ewels et al., 2016	https://multiqc.info/ ; RRID:SCR_014982
Subread 1.6.0	Liao et al., 2013	http://subread.sourceforge.net/ ; RRID:SCR_009803

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Picard tools 2.9.0	Broad Institute	https://broadinstitute.github.io/picard/ ; RRID:SCR_006525
Trimmomatic 0.36	Bolger et al., 2014	http://www.usadellab.org/cms/index.php?page=trimmomatic ; RRID:SCR_011848
PWMScan 1.1.1	Ambrosini et al., 2018	https://ccg.epfl.ch/pwmtools/
HOMER 4.9	Heinz et al., 2010	http://homer.ucsd.edu/homer/ ; RRID:SCR_010881
BiasAway 0.96	Worsley Hunt et al., 2014	https://github.com/wassermanlab/BiasAway
MEME suite 4.11	Bailey et al., 2009	http://meme-suite.org/ ; RRID:SCR_001783
Matrix-clustering	Castro-Mondragon et al., 2017	http://rsat.sb-roscoff.fr/matrix-clustering_form.cgi
DiffBind 2.8.0 Bioconductor package	Ross-Innes et al., 2012	https://bioconductor.org/packages/release/bioc/html/DiffBind.html ; RRID:SCR_012918
DESeq2 1.20.0 Bioconductor package	Love et al., 2014	https://bioconductor.org/packages/release/bioc/html/DESeq2.html ; RRID:SCR_015687
csaw 1.14.1 Bioconductor package	Lun and Smyth, 2016	https://bioconductor.org/packages/release/bioc/html/csaw.html
limma 3.36.5 Bioconductor package	Ritchie et al., 2015	https://bioconductor.org/packages/release/bioc/html/limma.html ; RRID:SCR_010943
precrc 0.10.1 R package	Saito and Rehmsmeier, 2017	http://takayasaito.github.io/precrc/
ChIPseeker 1.20.0 Bioconductor package	Yu et al., 2015	http://www.bioconductor.org/packages/release/bioc/html/ChIPseeker.html
chromVar 1.6.0 Bioconductor package	Schep et al., 2017	https://bioconductor.org/packages/release/bioc/html/chromVAR.html
Other		
HOCOMOCO v10 database	Kulakovskiy et al., 2013	http://hocomoco10.autosome.ru/ ; RRID:SCR_005409
JASPAR 2016 database	Mathelier et al., 2016	http://jaspar.genereg.net/api/ ; RRID:SCR_003030
TRRUST v2 database	Han et al., 2018	https://www.grnpedia.org/trrust/
ReMap 2015 database	Griffon et al., 2015	http://tagc.univ-mrs.fr/remap/
chromHMM data for the primary B cells	Kundaje et al., 2015	http://www.roadmapepigenomics.org/
GENCODE v29	Frankish et al., 2019	https://www.gencodegenes.org/

LEAD CONTACT AND MATERIALS AVAILABILITY

Further information and requests for resources should be directed to and will be fulfilled by the Lead Contact, Judith Zaugg (judith.zaugg@embl.de). This study did not generate new unique reagents.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Informed consent was obtained from all patients and the experiments on CLL samples was approved by the ethics committee of the medical faculty of the Heidelberg hospital under the number S-254/2016. Peripheral blood was taken from four CLL patients (2 male, 2 female, aged between 61 and 74; [Table S3](#)) and separated by Ficoll gradient (GE Healthcare), mononuclear cells were cryopreserved on liquid nitrogen. Variation related to gender were added as a confounding effect to the analysis and its effect not directly assessed. CLL samples were used in a paired design - i.e., CLL with treatment versus CLL with control treatment, thus gender and age are controlled for due to the design of the experiment.

METHOD DETAILS

Summary of publicly available data used in analyses

- (1) CLL: We obtained chronic lymphocytic leukemia data ([Rendeiro et al., 2016](#)) from the European Genome-phenome Archive (EGA: EGAD00001002110). The dataset consists of 88 ATAC-Seq samples from 55 patients (34 U-CLL, 50 M-CLL and 4 unclassified) as well as RNA-Seq samples for 10 patients (6 M-CLL and 4 U-CLL) ([Table S3](#)). However, except for the subsampling analysis, we here used only one sample per patient (the one with the suffix “_1”), resulting in a total of 52 samples

(25 U-CLL and 27 M-CLL). MPP/GMP: data described in this study was published as part of our earlier study (Rasmussen et al., 2019), and is deposited at GEO: GSE115972.

- (2) TF binding models: We used the mononucleotide position weight matrices (PWMs) data from the HOCOMOCO v10 database (Kulakovskiy et al., 2013) for 640 human and 422 mouse TF (Table S3) for all analyses (<http://hocomoco10.autosome.ru>), except for the comparison to the JASPAR database (see (5)).
- (3) ReMap: We used ChIP-seq data for hg19 from ReMap 2015 (Griffon et al., 2015) (<http://tagc.univ-mrs.fr/remap/index.php>).
- (4) TF-target interactions: We used data from TRRUST version 2 (Han et al., 2015), a manually curated database of human and mouse transcriptional regulatory networks (<http://www.grnpedia.org/trrust/>).
- (5) JASPAR 2016 (Mathelier et al., 2016): We used JASPAR 2016 core PFMs, parsed with the JASPAR RESTful API. (<http://jaspar.genereg.net/api/>).

Tools used

We used the workflow manager *Snakemake* (Köster and Rahmann, 2012) 5.0 for all described pipelines as well as the *Bioconda* project (Grüning et al., 2018) to maintain and handle software via the *Conda* package manager (<https://conda.io/en/latest/>). The ATAC-Seq preprocessing pipeline uses *bedtools* 2.26.0 (Quinlan and Hall, 2010), *samtools* 1.4 (Li et al., 2009), *Bowtie2* 2.3.0 (Langmead and Salzberg, 2012), *deepTools* 2.5.0 (Ramírez et al., 2014), *FastQC* 0.11.5 (Andrews, 2010), *GATK* 3.7 (McKenna et al., 2010), *MACS2* 2.1.1 (Zhang et al., 2008), *multiqc* 0.9 (Ewels et al., 2016), *Subread* 1.6.0 (Liao et al., 2013), *Picard* tools 2.9.0, *Trimmomatic* 0.36 (Bolger et al., 2014), *PWMScan* (Ambrosini et al., 2018) web interface (<https://ccg.epfl.ch/pwmttools/>) and standalone release 1.1.1 (<https://sourceforge.net/projects/pwmscan>) and R version 3.4. In addition, a number of Bioconductor (Huber et al., 2015) packages were used, the most important of which are described below. For a full list of used packages and their corresponding versions, see the Github repository at <https://git.embl.de/grp-zaugg/diffTF>. For the various additional analyses that we performed, we used the following tools: *HOMER* v4.9 (Heinz et al., 2010), *BiasAway* 0.96 (Worsley Hunt et al., 2014), *MEME* suite 4.11 (Bailey et al., 2009) and the Bioconductor *chromVAR* package (Schep et al., 2017).

ATAC-Seq processing

The *diffTF* pipeline requires properly processed and quality-controlled ATAC-Seq data, which can be obtained by any generic ATAC-seq processing. We used an in-house *Snakemake* pipeline that starts with raw fastq files and integrates multiple steps for quality control, adaptor trimming, alignment, as well as general and ATAC-Seq specific post-alignment filtering and processing steps (see workflow in Figure S1A). In the following we describe the pipeline using the default parameters. First, *FastQC* is used to assess the sequence quality. Foreign sequences from the Nextera Transposase agent are removed with *Trimmomatic*, using the parameters *ILLUMINACLIP:NexteraPE-PE.fa:1:30:4:1:true TRAILING:3 MINLEN:10*. Alignment is performed with *Bowtie2* with *-X 2000* (maximal fragment length), *-very-sensitive* and against hg19, followed by various cleaning steps (*Picard* tools *CleanSam*, *FixMateInformation*, *AddOrReplaceReadGroups*, and *ReorderSam*) and base quality recalibration using *GATK* with known variants taken from the provided *GATK* bundle for hg19 (SNPs: *dbSNP* version 138, Indels: *Mills_and_1000G_gold_standard.indels*). This allows us to detect and correct systematic errors made by the sequencer when estimating the quality score of each base call, thereby increasing data quality.

The pipeline then performs various cleaning and filtering steps: (1) removing mitochondrial reads and reads from non-assembled contigs or alternative haplotypes, (2) filtering reads with a mapping quality below a user-specified threshold (default = 10), (3) marking and removing duplicate reads with *Picard* tools, (4) adjusting read start sites as described previously (Buenrostro et al., 2013) (4 bp on the forward and 5 bp on the reverse strand) and (5) removing reads with insertions or deletions using *samtools* (Figure S3A).

Lastly, a GC bias diagnosis and correction using *deepTools* and Benjamini's method (Benjamini and Speed, 2012) is run for each sample. This helps to assess the severity of the GC bias in the data, namely that of DNA polymerases used for PCR-based amplifications during the library preparation, which usually results in artificially higher read counts for GC rich regions (Figures S3C and S3D).

The output of the pipeline can be obtained for the (1) original and (2) GC corrected data as well as on the level of (3) individual files (i.e., one file per replicate) or (4) pooled replicates (i.e., one file per sample). Downstream analyses for these four classes of files involve peak calling using *MACS2* with user-adjustable stringencies and parameters and removal of blacklisted regions. Finally the pipeline outputs summary statistics and additional files and plots (coverage files for visualization, transcription start site enrichment, sample-specific fragment length distributions (Figure S3B), library complexity measures, PCA (Figures S3E–S3J), sample correlations).

Description of the diffTF analysis workflow

In summary, *diffTF* calculates a differential TF activity between two or more conditions for each TF by comparing the distribution of fold-change differences across all binding sites of a TF to all binding sites from all other TFs. The algorithm is split into 7 steps as described in the following (see also workflow in Figures S1A and S1B):

1. Generating a consensus peak set

Our approach is based on analyzing read counts in peaks, which requires a consensus peak set across input samples. This can either be user-provided or it is generated from the sample-specific peaks. For the latter, consensus peaks are generated with the function *dba.peakset* in the *DiffBind* (Ross-Innes et al., 2012) Bioconductor package (using the parameter *minOverlap* to define the number of

samples within which a peak should be present). We then retain only peaks from genuine autosomes, thereby filtering sex chromosomes, non-assembled contigs as well as alternative haplotypes. The consensus peak set is finally sorted by coordinate to speed up subsequent computations.

2. Scanning of TF binding sites

For each TF of interest, *diffTF* needs a set of TF binding sites (TFBS). To generate them, we used the *HOCOMOCO* database, which provides TF binding models (PWMs) that are collected from numerous ChIP-Seq experiments for 640 TFs in human and 422 in mouse (Table S2). We used these PWMs to scan the *hg19* genome using the *PWMscan* web interface and the standalone version to obtain all putative binding sites for each TF (cutoff p value - 0.00001, background base composition - 0.29;0.21;0.21;0.29). We finally sort the TFBS for each TF by coordinates. For user convenience, we provide this sorted list for both human and mouse in the Git repository. However, *diffTF* is not limited to any specific database and any tool can be used to predict TFBS.

3. Differential analysis for the consensus peakset

To calculate the fold-change between the two conditions across each peak, we first obtain the counts for the consensus peakset for each sample using *featureCounts* from the *Subread* package with the options -p -B -d 0 -D 2000 -C -Q 10 -O -s 0. We then employ *DESeq2* (Love et al., 2014) for count normalization using a cyclic loess approach as implemented in the *normOffsets* function of *csaw* (Lun and Smyth, 2016) from Bioconductor to alleviate potential systematic biases between the samples. Importantly, unrestricted design formulas can be incorporated into *diffTF*, making it therefore a very flexible approach if additional co-variates should be incorporated to increase statistical power. In addition to the two conditions comparison as described above (e.g., mutated versus unmutated, wild-type versus mutant, young versus old), *diffTF* can also incorporate design formulas for which the predictor variable is continuously-valued rather than binary to analyze time-course data (e.g., age, differentiation time, etc). The reported log2 fold-change is then per unit of change of that variable. That is, TFs with a negative differential *TF activity* have a negative slope per unit of change of that variable, while TFs with a positive *TF activity* have a positive one. To check the normalization results, we generate four types of diagnostic plots: (1) MA plots (Figure S2I), (2) density plots of normalized and non-normalized counts (see Figure S2H for the normalized counts), (3) mean-average plots (average of the log-transformed counts versus the fold-change per peak) for each of the sample pairs and (4) mean SD plots (row standard deviations versus row means).

4. Signal extraction for each TFBS

We then filter for TFBS that overlap with the consensus peak set only while allowing for multiple TFBS per peak (using the program *bedtools intersect* with options -wa -wb). Each TFBS is then extended by 100 bp (user-adjustable) in both directions (motif extension), followed by extraction of read counts for each sample (*featureCounts*, see above for the parameters used).

5. Calculation of accessibility fold-change for each TFBS

To avoid biases and dependencies based on TFBS clustering within peaks, we then select the TFBS per TF per peak with the highest average read count across all samples. On this set we perform a differential accessibility analysis using a standard *limma* (Ritchie et al., 2015) workflow employing the functions *lmFit* and *eBayes* independently for each TF for the same design formula as used in step 3 (that is, either a model contrasting two conditions or a continuously-valued model). For count normalization, we use the consensus peak-derived normalization factors from step 3. As a result we obtain a log2 fold-change value for each selected TFBS per TF per peak, and in addition we provide various diagnostic plots for each TF (same as in step 3, with additional ECDF and density plots of the log2FC values, Figures S2H–S2J).

6. GC binning and calculation of differential TF activity values (see also Figure S1B)

(a) *GC binning.* To reduce biological biases based on differential effects depending on GC content of the local environment of a TF, we first obtain the GC content of each extended TFBS with *bedtools nuc*. We then group TFBS into 10 bins (user-adjustable) based on the GC content of their extended TFBS (e.g., +/-100bp).

(b) *Estimation of the differential TF activity.* For each bin containing more than 20 TFBS for a given TF, we compare the TF-specific distribution of log2 fold-changes against a background of log2 fold-changes of all TFBS from all other TFs of the same GC bin. We then define the difference in means as mean difference for each bin. Finally, for each TF, the differential *TF activity* is calculated as the weighted arithmetic mean across all mean difference values for the bins with sufficient data, weighing each value from each bin by the fraction of TFBS it contains in that TF such that all weights sum to 1.

7. Estimation of significance for differential activity for each TF

The significance and magnitude of the TF-specific *TF activity* values can be difficult to interpret. To assess their statistical significance, we employ a permutation approach to derive empirical p values (hereafter called the empirical approach). We rerun steps 3–6 for a total of 1000 times (user-adjustable) with permuted input data (permutation of condition labels) and then calculate an empirical two-sided p value per TF by comparing the real value with the distribution from the permutations and calculating the proportion of sampled permutations for which the absolute differential *TF activity* is larger (Figure S2A). To validate our procedure, we also plot the density of the *TF activity* across all permutations and compare it to the real data, which expectedly gives a distribution that is tightly centered around 0 for the permuted data while the real data show a shift as well as heavier tails (Figure S2B). We finally perform multiple testing correction using Benjamini-Hochberg (Benjamini and Hochberg, 1995).

If the number of samples is small or only very few samples come from one condition, the number of possible permutations is also small, therefore making it difficult for the permutation-based approach to accurately assess significance. In such a case, we offer a modified version of the described pipeline that we call the analytical approach: First, we use *DESeq2* instead of *limma* to calculate

fold-changes throughout the pipeline for the TF-specific steps, which is slower but better suited for a small number of samples. Second, instead of employing permutations, we run a modified procedure from step 7 onward to assess the statistical significance of the differential *TF activity*: To reduce the dependency of the p value on the sample size (since the number of TFBS can range between a few dozen and multiple tens of thousands depending on the TF), we first perform a Welch Two Sample t test for each bin and calculate the overall significance by converting the resulting T-scores to z-scores, which allows to summarize them across the bins and convert them to one p value per TF (Figures S2D and S2E). Because we calculate the weighted mean out of the individual T-scores and therefore end up having a weighted sampling distribution of the mean, we however need to estimate its expected variance.

(a) Estimation of the variance

To estimate the variance of the weighted mean T-scores for each TF we use the following formula (see https://en.wikipedia.org/wiki/Propagation_of_uncertainty#Linear_combinations) in which w_i and x_i represent the weight and T-statistic value from the t test of a particular bin i , $Cov(x_i, x_j)$ the covariance of two particular bins i and j , and $var(x_i)$ the variance of the T-statistic value for x_i :

$$var(\underline{x}) = \sum_{i=1}^n var\left(\frac{w_i}{\sum_{k=1}^n w_k} x_i\right) + 2 \underbrace{\sum_{i \geq 1} \sum_{j > i}^{j \leq n} w_i w_j Cov(x_i, x_j)}_{\text{(weights sum up to 1)}}$$

$$var(\underline{x}) = \sum_{i=1}^n var(w_i * x_i) + k \quad (\text{if all values are scaled by a constant, the variance is scaled by its square})$$

$$var(\underline{x}) = \sum_{i=1}^n w_i^2 var(x_i) + k$$

This is done to prevent a systematic misestimation of the variance of the x_i terms by assuming it to be 1 for all bins. To properly estimate $var(x)$, we employ a bootstrap approach using the *boot* library in R with a user-adjustable number of bootstrap replicates (default 10,000), with resampling the bin-specific data and then performing the t test against the full sample as described above. We then calculate the variance $var(x_i)$ of the bootstrapped T-scores for each bin. Since the T-scores across the bins are not necessarily independent, we also correct their variance using their pairwise covariance, which is estimated with the term $Cov(x_i, x_j)$ using the bootstrap values for each bin. Across different analyses and datasets, we generally observed that covariances are predominantly negligible, but for some TFs and bins, they differ significantly from 0 and therefore it is important to incorporate them. A detailed analysis of the covariances $Cov(x_i, x_j)$ between pairs of bins of the TF-specific distribution of log2 fold-changes against a background of log2 fold-changes from a randomly selected set of TFBS of the same GC bin (see step 6 and the formula in step 7a) revealed that covariances are negligible in almost all cases. From the 9,099 covariances across all TF and pairs of bins, only 19 cases exceeded an absolute value of 0.05 and only one 0.2, while over 91% had absolute values of smaller than 0.02 (data not shown). We verified the validity of this procedure with a simulation script (available upon request) that shows that p values are well behaved for random data (Figure S2G).

(b) Calculation of one p value per TF

To obtain a p value for each TF, we centralize the distribution of the weighted T scores across all TF by subtracting its mean, which we obtain from the maximum likelihood estimate of the distribution mean using the *locfdr* function from the *locfdr* package in R. We then calculate p values out of z-scores based on the TF-specific variance calculated above. We finally perform multiple testing correction using Benjamini-Hochberg method (Benjamini and HochbergY, 1995).

Validation of the analytical approach

We used the GMP/MPP data to verify the validity of our analytical approach by comparing the real signal from the GMP versus MPP comparison with an array of additional analyses: We first run *diffTF* within GMP and MPP, respectively, and repeated this for all possible pairwise combinations (2 versus 2). In addition, we run two types of controls: One regular *diffTF* analyses of GMP versus MPP in which we shuffled the TFBS before the binning and calculation of statistical significance and one in-silico simulation based on random data that mimics precisely the *diffTF* workflow while incorporating some of the GMP versus MPP specific parameters (see STAR Methods). Overall, we found that comparing GMP versus MPP gives the strongest signal, while comparing only within GMP or MPP cells still yields signal albeit on a much smaller scale, while the two types of control indeed showed no signal (Figure S2G). The remaining signal within cell types is a tribute to the high sensitivity of *diffTF* to pick up small signal and reflects the inter-individual differences between biological replicates. Lastly, we compared TF activity and p values from the analytical with the empirical approach, the latter of which was run with all possible 70 permutations that are possible for a 4 versus 4 analysis. The TF activity values were almost identical (Figure S6C, top), which is not surprising given that the only methodological difference is the usage of DESeq2 and limma for calculating log2 fold-changes for the empirical and analytical approach, respectively, which also highlights technical robustness. The ranks of the adjusted p values also correlated significantly (Figure S6C, bottom). We therefore recommend for the analytical approach to perform at least one within-sample analysis to obtain an idea of the expected signal based on inter-individual variation of biological replicates.

Guidance on the number of permutations and whether to use the analytical or permutation-based approach

As explained before, *diffTF* offers two procedures to calculate significance, and we want to give some guidance what the significance in each procedure means, when to use which, and how to adjust the parameters. We note again that the calculation of the TF activity is not affected by this choice, they are practically identical between the two approaches, while p values may differ although they are also correlated (Figure S6C).

(1) Permutation-based approach

In this approach, the resulting significance value captures the significance of the effect size (that is, the *TF activity*) for the real data as compared to permuted one. Importantly, this significance measure appears independent of the number of binding sites per TF ($r = 0.026$, Spearman, see also Figure S2C), and should be used when the number of samples is large enough to perform a reasonable number of permutations. The range of p values will be determined by the number of permutations (smallest p value equals to $1/\{\text{\# of permutations}\}$) and we recommend at least 1000 permutations (as we have done for the CLL dataset). The number of possible permutations can be calculated with the binomial coefficient (n/k), with n being the total number of samples across the two conditions and k either one of the number of samples per condition. For example, when comparing 5 GMP against 4 MPP samples, there are $(9/5) = (9/4) = 126$ possible permutations. The minimum raw p value $p_{\min} > 0$ is then $(1/126) \sim 0.008$. This approach is computationally more expensive than the analytical one.

(2) Analytical approach

Here, the resulting significance value is calculated by performing t tests for each TF and GC bin (see STAR Methods above). Given that both the p value and to a lesser extend also the T statistic depends on the sample size (i.e., the number of binding sites n), we checked whether this dependence would explain much of the signal. However, for both CLL and GMP/MPP, we found highly significant TFs with lower number of binding sites despite the correlations of p value and number binding sites ($r = -0.158$ and $r = -0.409$, respectively, Spearman), thus showing that the biological signal captured by *diffTF* surpasses the theoretical bias (Figure S2F).

If the number of samples is too small to allow a reasonable number of permutations in the permutation-based approach or computation time is important, this approach can be used. The number of bootstraps for the variance estimation of the T scores should be set to a high value, and we recommend values of at least 1,000 or even 10,000. Computation time, however, increases with the number of bootstraps. For the CLL data, the estimated bootstrap variance approaches the expected value 1 very quickly even for bins with low numbers of TFBS (the minimum is set to 20).

Classification of TFs into activator and repressors

First, we tried to classify TFs based on literature-mining, using *TRRUST*. However, we found that most TFs were classified almost equally often as activator and repressor, which makes it very difficult to determine an overall function for each TF (see Figure S5C). This is likely due to regulatory interactions being affected by multiple factors such as cell type, study conditions, highly studied TFs and different experimental conditions. Therefore, we decided to develop a cell-type specific data-driven approach to classify each TF into putative activator, repressor or undetermined, based on the overall effect on their putative TFBS. Our classifier is based on the assumption that increasing the level of an activating TF increases chromatin accessibility at its target sites while increasing the level of a repressing TF decreases it. For this, we first quantile normalized the RNA-Seq count data in order to minimize the effects of outlier values that may have a large influence on the resulting correlations otherwise. We then calculated the Pearson correlation coefficients between the expression level of each TF and the ATAC-Seq signal of each putative TFBS across all individuals (see Figure S5B for an example). If the median of the resulting correlation coefficients was sufficiently positive we consider it an activator, if sufficiently negative as a repressor, and if it was not significantly different from the background we call it undetermined (see below for what sufficiently refers to). In detail, for each TF, we calculated the correlation median across the foreground (i.e., all peaks with predicted TFBS for the particular TF) and the background (i.e., all other peaks - peaks without predicted TFBS for the particular TF). The latter was used to estimate the noise level for the median correlations for each TF. In particular, to distinguish real correlations from noise (i.e., activator/repressor from undetermined) we used particular percentiles of the background distribution across the background values for all TF as a threshold for activators and repressors, respectively. In order to not be dependent on one particular parameterization with respect to the resulting classification, we provide different variants corresponding to different stringency thresholds (10/90%, 5/95%, 1/99%, 0.1/99.9%, 0.01/0.999%) and give the classification for each of these, thereby providing flexibility for subsequent analyses and the required level of stringency. Lastly, as additional filter and in addition to the percentile cutoff, we also quantified whether the foreground and background distribution are significantly different from another using a one-sided Wilcoxon rank sum test. For TFs that were classified as either repressor or activator but for which the raw p value of the Wilcoxon rank sum test was not significant, we changed their classification to undetermined, thereby removing TF classifications with weak support. The resulting classification of activators and repressors for positive and negative medians, respectively, was used to assess the level of coherence based on TF expression and their *TF activity* calculated by *diffTF*. We also summarized the fraction of the activators/repressors in certain TF cluster from RSAT clustering (see Figure S5A for an example). Notably, most TF families as defined above with the RSAT clusters, comprised both, predicted activators and repressors, indicating that the molecular function changes even within a TF family.

CLL ATAC-Seq processing

We run the ATAC-Seq pipeline on the 52 CLL samples as described above with the indicated default parameter values. The distribution of the number of the fraction of retained and total number of reads across all samples throughout the course of the pipeline showed the typical pattern, namely that reads were filtered mainly because they were of either of mitochondrial origin or duplicates (Figure S3A). The fragment length distributions also showed the typical and expected pattern, with peaks at expected lengths (e.g., mono- and dinucleotides) (Figure S3B). Lastly, they showed that read counts were partially driven by the GC content (see Figure S3C for a typical profile from a randomly chosen sample), which we eliminated using the GC correction as described. The diagnostic plots after the GC correction then expectedly indicated no GC dependency (see Figures S3C and S3D). We also performed a PCA analysis of the consensus peak regions (Figures S3E–S3J).

Running diffTF on the CLL data

We used 52 samples out of 88 available ATAC-Seq samples from the CLL data. We excluded 4 samples because their mutational status was unknown. Out of 84 remaining samples, we used only one replicate per sample (“_1”), which resulted in a total of 52 samples (25 U-CLL and 27 M-CLL, see also Table S3) in order to not bias the analysis by overrepresenting particular individuals due to the varying number of replicates per sample. As discussed above we used the GC-corrected BAM files hereafter. A PCA analysis of these peak regions showed a clear separation of the samples based on their IgHV mutation status, but not on any other potentially confounding factor like batch, IGVH homology, gender, the patient age at data collection or the patient age when diagnosed (Figures S3E–S3J). We generated a consensus peakset as described above using 5 samples as the minimum overlap. This retained 48,065 consensus peaks which were used subsequently. As motif extension, we added 100 bp on each side of the predicted *in silico* TFBS. As parameters for the peaks and TF-specific differential analysis, we used the design formula: “~ batch + mutationStatus” and *fitType* = ‘local’. Here, *batch* and *mutationStatus* are metadata that are available for each sample and that refer to sample batch and its mutation status (either mutated or unmutated), respectively. Based on the discussion in step 6, we decided to use 10 bins for our GC binning approach (that is, ranging from 0%–10% GC content up to 90%–100%).

Comparison of diffTF results based on predicted versus ChIP-Seq validated TFBS

We used TF binding data gathered by *ReMap* from human ChIP-seq experiments and intersected them with *in silico* predicted TFBS using TF binding models from the HOCOMOCO database. We then split the 157 common TFs by whether or not they overlapped with the *ReMap* ChIP-seq data and rerun the *diffTF* pipeline. Finally, we correlated these two groups with and without ChIP-seq validation using Pearson correlation. The correlation between the TFBS that were present versus not present in *ReMap* was 0.54 ($p = 3.5e-10$) (Figures S4A and S4B).

Assessing impact of TF motif scanning parameters on diffTF results

For the purpose of assessing the effect of the genome-wide TFBS scanning on *diffTF*, we varied the *p* value cutoff in *PWMScan* (0.00005, 0.00001, and 0.000001) using the default background base composition of 0.29;0.21;0.21;0.29 as well as varying the background base composition (0.27;0.23;0.23;0.27) (Figures S4C and S4D). The former resembles the human genome nucleotide composition, the latter the base composition from the ATAC-Seq peak regions only. For this, we generated the FASTA file of the consensus peak set with *bedtools getfasta* and then calculated the background nucleotide composition with *fasta-get-markov* from the *MEME* suite. PWMs from HOCOMOCO and PFMs from JASPAR were converted into integer log likelihoods using *pwm_convert* (setting *-f* as “real” and “jaspar” for PWM and PFM, respectively). Scanning was performed using the script *pwmBowtieWrapper*.

diffTF robustness analysis with respect to internal parameters for the CLL data

As discussed above, we verified the robustness against internal parameters such as the number of permutations or the number of bootstraps (for the analytical significance calculations) and found that *diffTF* results are not dependent on these. Most importantly, we also systematically varied the default motif extension size from the default 100 bp to 0, 50, 200, 400, and 600 bp to investigate the effect of the extended TFBS on the *diffTF* results (Figures S4E and S4F). In summary, the *TF activity* differences were very similar, with correlations above 0.9 and 0.86 for motif extension sizes from 50–600 for Pearson and Spearman, respectively. Only an extension size of 0 showed weaker correlation, albeit the values were also surpassing 0.76 and 0.71, respectively. For almost all TFs, the *TF activity* difference estimates were consistent across motif extension sizes, with only a few TFs changing directions from positive to negative values or vice versa. For large extension sizes, the magnitude of the differential *TF activity* decreased overall. Running *diffTF* with no extension size or too large values (> 500), therefore, seems inappropriate as it either excludes too much signal directly adjacent to the predicted TFBS or the resulting extended TFBS are becoming too wide.

Raw and adjusted *p* values had more differences when varying the extension size, with correlations above 0.53 and 0.58 for extension sizes from 50–600 for Pearson and Spearman, respectively. Not using any motif extension showed again weaker correlation with values at least being 0.28 and 0.35, respectively. In conclusion, the raw and adjusted *p* value comparison across extension sizes identified three clusters of TF: a small number of TFs gaining or losing significance with increasing extension size, respectively, and the majority of TFs being invariant for it. 67% of all TFs that we identified as being significant remained significant throughout the majority across all extension sizes, while almost 95% were significant for at least one other extension size.

Assessing dependence of *diffTF* results on differential signal strength for CLL data

In order to assess the effect of differentially accessible (DA) peaks on *diffTF*, we generated a peak set excluding DA peaks. For this, we used *DiffBind* (with design $\sim \text{batch_number} + \text{Condition}$) and identified 389 and 3569 DA peaks for the mutated and unmutated condition, respectively, which we then excluded from the original peak set. We then run *diffTF* with standard parameters (described above) and correlated the results (see Figure 2F).

GO terms associated with CLL progression

We obtained ENSEMBL gene IDs from GENCODE v29 (Frankish et al., 2019) with the *org.Hs.eg.db* Bioconductor package v3.8.2 and linked them to the related gene ontology (GO) terms for all biological processes. We then used *precrec* (Saito and Rehmsmeier, 2017) R package v0.10.1 to calculate ROC curves for each GO term based on all significant TFs from *diffTF* (adjusted p value < 0.1). For filtering, we kept only GO terms with AUC > 0.6 and those with at least 9 but less than 90 TFs to remove GO terms associated with DNA-binding, which are expected to be enriched for TFs.

Assessment of the power of *diffTF* with respect to sample size and sequencing depth for CLL data

To test the power of our approach and the dependence on the (i) read depth (coverage) of the BAM files and (ii) number of samples for each of the two conditions, we performed a subsampling procedure in which we varied both read depth and sample size. We used the full 84 sample CLL dataset for this in order to maximize the number of samples. To vary the read depth, we used the original, non-GC corrected CLL data that is produced as part of the output of our ATAC-Seq pipeline and then randomly generated downsampled BAM files using Picard *DownsampleSam* with random seeds and PROBABILITY = k, with k ranging from 0.75, 0.50, 0.25, 0.125, 0.06, 0.02 to 0.01. For each of these fractions, we (1) determined the median number of reads these fractions correspond to across all samples, thereby giving an estimate of the required number of processed reads to produce a particular accuracy and (2) run the *diffTF* pipeline with the addition of also varying the number of samples per condition while maintaining the original ratio of the two conditions (around 60% to 40%). The latter we varied from the full data 34+50 down to 30+45, 25+37, 20+30, 15+23, 10+15, and down from 9+14 in steps of 1 to 3+5, with the first and second number denoting the number of distinct samples with the condition unmutated and mutated, respectively. For each of these cases, we in addition performed 50 repetitions to minimize sampling noise. We then compared the values from the full dataset with each of the two-way subsampled data and evaluated the results by assessing the fraction of TFs that show the same direction of change (that is, either positive or negative) for the differential TF activity as in the full data. We split all TFs into three equally sized bins using the 33% and 66% quantile threshold with respect to the absolute differential TF activity in order to differentiate between TFs with low, medium and high signal, respectively (Figure S4G).

RNA-seq processing for CLL data and classification into activators and repressors

We used all ten available RNA-Seq samples from (Rendeiro et al., 2016) as well as available metadata (e.g., age, sex and condition). Initial quality control and adaptor trimming (using *ILLUMINACLIP:Truseq-2.fa: 1:30:4:5:true TRAILING:3 MINLEN:20* instead) was performed as described in the ATAC-Seq pipeline. We then aligned the samples to hg19 using STAR (Dobin et al., 2013) with the parameters *-outFilterMultimapNmax 2-quantMode GeneCounts* using the Gencode (Harrow et al., 2012) v29 annotation. We filtered 2 out of 10 samples due to data quality that we identified via PCA (data not shown; see also Table S3). We finally employed *DESeq2* with the design formula " $\sim \text{condition}$," filtered genes with either (1) less than 5 reads on average in either one of the conditions or (2) a median expression value of 0, and identified those genes corresponding to the TFs from *HOCOMOCO*. From the initial list of 640 TFs, this excluded 270 TFs that were not-expressed according to these criteria, and 370 TFs that we labeled as being expressed and that we used for the classification into activator or repressor. For the subsequent classification in *diffTF*, we chose to use the 5/95% cutoff. See also Figures S5B and S5C for details.

Correlation of TF activity with target genes expression for CLL data

For the CLL target analysis we used the same set of predicted TFBS using PWMscan as with *diffTF* and annotated each TFBS to the closest gene using the *ChIPseeker* (Yu et al., 2015) package in Bioconductor (Huber et al., 2015) using the hg19 annotation. TFBS located -2,000/+500 bp from the TSS were annotated as promoter TFBS and used for the analysis. Using expression data from RNA-seq as described above, we then calculated the median log2 fold-changes of the target genes of a given TF that were defined as having at least one TFBS of interest in the promoter regions. For the final representation in Figure 4C, we used only TFs that had more than 200 and less than 1,500 unique target genes.

Footprinting analysis for CLL data

For the CLL footprinting analysis, we selected all 25 U-CLL samples and the 25 samples for M-CLL with the highest read counts, downsampled all to the sample with the lowest read count (~14.5 million) and merged all U-CLL and M-CLL samples. For all TFs that are expressed according to the RNA-Seq data, we then took the same set of TFBS as for *diffTF*, without TFBS that were overlapping between activators and repressors, and run the *dnase_to_javareview.py* script from the *pyDNase* library (Piper et al., 2015) for each condition to obtain base-specific Tn5 insertions matrices for each TFBS. These values were then normalized to the amount of reads in the consensus peakset, which was calculated using *featureCounts* from the *Subread* package with the parameters *-p -B -d 0 -D 2000 -C -Q 10 -O -s 0*. Finally, we removed peaks in blacklisted regions and potential artifacts (> 1000 counts). For generating

the genomic background for the openness, we binned the consensus peakset with a bin size of 200 bp and randomly selected 10,000 regions for which we then ran again the *dnase_to_javatreeview.py* script followed by subsequent normalization. We then performed a PCA analysis and visualized the scatterplot of the first two PC components that collectively contain ~36% of the overall variance (Figure 3D). For this, we normalized the data as follows: For each TF and base pair, we scaled the value for the Tn5 insertions by dividing it with the TF-specific average number of Tn5 insertions (row mean). For the footprint plots summarizing each quadrant of the PCA (Figure 3D), we used the same normalized the data for all TFs in the specific quadrant.

Chromatin state enrichment in CLL data

For each TF deemed significant with *diffTF* that was also expressed according to the RNA-Seq data, we intersected all TFBS that we obtained with the genome-wide *PWMscan* scanning (see above for details) with the expanded 18-state model from the *chromHMM* (Ernst and Kellis, 2012) data of the primary B cells (Kundaje et al., 2015). For each TF, we then calculated the fraction of its TFBS overlapping each state. We then grouped the TFs into activators and repressors, visualized their respective distributions with box-plots, and statistically assessed the differences between activators and repressors using a Wilcoxon rank sum test.

CLL data treated with Ibrutinib

Peripheral blood was taken from 4 CLL patients and separated by Ficoll gradient (GE Healthcare), mononuclear cells were cryopreserved on liquid nitrogen. Samples were later thawed from frozen as previously described (Dietrich et al., 2018) and MACS sorted for CD19 positive cells (Milteny autoMACS®). The cells were resuspended in RPMI (GIBCO, Cat.No. 21875-034), with the addition of 2mM glutamine (GIBCO, Cat.No. 25030-24), 1% Pen/Strep (GIBCO, Cat.No. 15140-122) and 10% pooled, heat-inactivated and sterile filtered human type AB male off the clot serum (PAN Biotech, Cat.No. P40-2701, Lot.No:P-020317). 5ml of cell suspension was cultured in 6-well plates (Greiner Bio-One Cat.No. 657160). To prepare the treatments, Ibrutinib (Selleckchem, Cat.No. S2680) was dissolved in Dimethyl sulfoxide (DMSO; SERVA, Cat.No. 20385) and stored at –20°C. After thawing, Ibrutinib was prediluted in DMSO and was added to the plates. Control wells were treated with DMSO in the same concentration as with Ibrutinib treatment. In both treatment and control, the final DMSO concentration was 0.2%. Cells were incubated at 37°C and 5% CO₂ for 6 hours with or without 500nM ibrutinib. The final cell concentration was 2x10⁶ cells/ml. After treatment, cell viability and purity was assessed using FACS. All samples had a viability over 90% and over 95% of CD19+/CD5+/CD3- cells.

ATAC-seq libraries generation of the CLL treated with Ibrutinib

ATAC-seq libraries were generated as described previously (Buenrostro et al., 2013). Cell preparation and transposition was performed according to the protocol, starting with 5x10⁴ cells per sample. Purified DNA was stored at –20°C until library preparation was performed. To generate multiplexed libraries, the transposed DNA was initially amplified for 5x PCR cycles using 2.5 µL each of 25 µM PCR Primer 1 and 2.5 µL of 25 µM Barcoded PCR Primer 2 (included in the Nextera index kit, Illumina, San Diego, CA, USA), 25 µL of NEBNext High-Fidelity 2x PCR Master Mix (New England Biolabs, Boston, Massachusetts) in a total volume of 50 µL. 5 µL of the amplified DNA was used to determine the appropriate number of additional PCR cycles using qPCR. Additional number of cycles was calculated through the plotting of the linear Rn versus cycle, and corresponds to one-third of the maximum fluorescent intensity. Finally, amplification was performed on the remaining 45 µL of the PCR reaction using the optimal number of cycles determined for each library by qPCR (max. 13 cycles in total). The amplified fragments were purified with two rounds of SPRI bead clean-up (1.4x). The size distribution of the libraries was assessed on Bioanalyzer with a DNA High Sensitivity kit (Agilent Technologies, Santa Clara, CA), concentration was measured with Qubit® DNA High Sensitivity kit in Qubit® 2.0 Fluorometer (Life Technologies, Carlsbad, CA). Sequencing was performed on NextSeq 500 (Illumina, San Diego, CA, USA) using 75bp paired-end sequencing, generating ~450 million paired-reads per run, with an average of 55 million reads per sample.

RNA-seq library generation for the CLL dataset treated with Ibrutinib

RNA was isolated using the miRNeasy Mini Kit (QIAGEN, Cat.No. 217004), starting with 1x10⁷ cells per sample. Cells were lysed in QIAzol Lysis reagent and homogenized using QIAshredder (QIAGEN, Cat.No. 79654), homogenized cell lysates were stored at –80°C until RNA extraction. RNA extraction was performed according to miRNeasy protocol and purified RNA was stored at –80°C until further processing. RNA integrity was checked using the RNA Nano 6000 Assay Kit of the Bioanalyzer 2100 system (Agilent Technologies, Santa Clara, CA), and concentration was measured with Qubit® RNA Assay Kit in Qubit® 2.0 Fluorometer (Life Technologies, Carlsbad, CA). Stranded mRNA-Seq libraries were prepared from 250ng of total RNA using the Illumina TruSeq RNA Sample Preparation v2 Kit (Illumina, San Diego, CA, USA) implemented on the liquid handling robot Beckman FXP2. Obtained libraries that passed the QC step, which was assessed on the Agilent Bioanalyzer system, were pooled in equimolar amounts. 1.8 pM solution of each pool of libraries was loaded on the Illumina sequencer NextSeq 500 High output and sequenced uni-directionally, generating ~450 million reads per run, each 85 bases long. RNA-seq processing of the data was done analogous as described above for the original CLL data, using DESeq2 with the design formula “~ treatment + mutational status.”

Comparing diffTF results with HOMER for CLL data

We also assessed the detection power of *diffTF* compared to a standard motif enrichment analysis (Figures S7A and S7B). For this, we generated a set of background sequences for each set of DA peaks in the two conditions, considering the length and GC bias as

features for the background generation. We then used *HOMER* (Heinz et al., 2010) to perform an enrichment analysis of motifs from *HOCOMOCO* on the DA peaks of each condition as foreground and respective two-fold *BiasAway* (Worsley Hunt et al., 2014) generated-sequences as background. We then correlated the *diffTF* results with the percentage of foreground sequences enriched for the given motif (*HOMER* output).

Strikingly, no enriched motifs were found in M-CLL with *HOMER*, while the few discovered in U-CLL correlated significantly with differential *TF activity* as computed by *diffTF* (Figures S7A and S7B). While we cannot exclude the possibility that some of the TFs identified by *diffTF* are false positives, this analysis, together with the observation that *diffTF* recapitulated many of the known TFs that distinguish U-CLL from M-CLL (see Figure 2A), highlights the power of *diffTF* to capture more signal than standard motif enrichment approaches.

Comparison with similar tools

We first compared *diffTF* with a more traditional TF motif analysis such as *HOMER* (Heinz et al., 2010), which looks at motif enrichment in a set of differentially accessible peaks. Strikingly, no enriched motifs were found in M-CLL with *HOMER*, while the few discovered in U-CLL correlated significantly with differential *TF activity* as computed by *diffTF* (Figures S7A and S7B; see also STAR Methods). While we cannot exclude the possibility that some of the TFs identified by *diffTF* are false positives, this analysis, together with the observation that *diffTF* recapitulated many of the known TFs that distinguish U-CLL from M-CLL (see Figure 2A), highlights the power of *diffTF* to capture more signal than standard motif enrichment approaches.

To compare *diffTF* with an approach that is also based on TF activities we chose *chromVAR* and *BaGFoot* (Baek et al., 2017). We were unfortunately unable to run and adopt the *BagFoot* workflow for our CLL data due to missing example files and incomplete documentation. For *chromVAR*, which was originally developed for analyzing single cell ATAC-seq data, the results correlated very well overall, with correlation coefficients between 0.75 and 0.93 (Pearson) and 0.69–0.88 (Spearman), depending on the set of TFs (i.e., all TFs or only those deemed significant by *diffTF*, therefore predominantly removing TFs with low signal) and whether *chromVAR* deviations or deviation scores were compared against (Figures S7C and S7D). Differences likely arise due to distinct methodological divergences such as comparing fold-changes for peaks (*chromVAR*) versus binding sites (*diffTF*) or whether (*diffTF*) or not (*chromVAR*) to normalize the TF-specific effect against the mean effect across all TFs (see Figures S7E and S7F and the more detailed assessment below). However, *diffTF* goes one step beyond the currently available methods by classifying TFs based on their mode of regulation – activator or repressor, thus providing important additional insights into their molecular function.

For the comparison of *diffTF* with *chromVAR* (Figures S7C–S7H) we generally mimicked our *diffTF* analysis in the *chromVAR* framework as much as possible. First, we imported all TFBS that overlapped with our peaks and that were used in *diffTF* using the *getAnnotations* function in *chromVAR*. We then used the same consensus peakset as input as for *diffTF* (without resizing the peaks to maximize compatibility). We then derived fragment counts in paired-end mode for the same BAM files using the function *getCounts*, followed by GC correction using *addGCBias*, peak filtering using *filterPeaks* using the default arguments. To compute the expectations, we used the CLL group assignments into U-CLL and M-CLL while also normalizing counts, which yields within each group the average fraction of reads per peak in each sample. After calculating the deviations and deviation scores, we summarized the results per TF by calculating the difference of the means between U-CLL and M-CLL samples separately for both of them. Finally, we correlated the *chromVAR* results with the differential *TF activity* from *diffTF*.

Generally, as stated and quantified above, *diffTF* and *chromVAR* correlate significantly (Pearson; Figures S7C and S7D) and agree for most TFs, but there are some at first sight discrepancies for TFs located in the second quadrant. They can likely be explained by the following methodological differences:

1. *chromVAR* compares only peaks and not individual TFBS, while *diffTF* computes a log2 fold-change for each TFBS. While peak log2 fold-change and corresponding TFBS log2 fold-change overall correlate highly (Pearson 0.91), TF-specific differences are in the range of –0.038 to 0.026 for expressed TFs in log2 fold-change units, which is considerable (47.5 and 26%, respectively) given that the *diffTF TF activity* score is in the same scale and ranges from –0.08 to 0.10 (Figures S7E and S7F for two specific examples and Figure S7G for a summary across TF). Overall, we observe that the mean log2 fold-change across all TFBS is 0.026, while for peaks, it is only 0.019.
2. Unlike *chromVAR*, *diffTF* compares each TF against the mean effect across all TF and therefore uses a relative rather than an absolute value. This explains most of the shift from TFs in Figure S7C in the left upper quadrant because overall, the mean log2 fold-change across all TFBS is slightly skewed to the positive. Thus, most *chromVAR* deviation values are consequently mostly positive (424 out of 640, ~66.3%), while *diffTF TF activity* values are mostly negative (420 out of 640, ~65.6%).
3. *chromVAR* never directly compares the two conditions with one another to compute log2 fold-changes, but only uses the conditions to compute a deviation that is based on the condition-specific expectation. Consequently, it provides a deviation value for each sample, which then have to be summarized accordingly. The recommended and used approach is to compute the mean deviation within each condition, the difference of which then mimics the differential *TF activity* as used in *diffTF*. However, the computation of the mean can be prone to outliers (Figure S7H). *diffTF* does not directly suffer from this issue as it compares the two groups with one another to derive a log2 fold-change.

From all expressed TFs for which *chromVAR* and *diffTF* differ in their predicted direction of change, we checked in our literature review for the TFs that were deemed significant in *diffTF* which CLL condition they have been previously been associated with (Table S2). Only 8 TFs are located in quadrant 2 (GCM1, NR4A3, NR1H2, NR1D1, PPARD, ESR1.A, NF2L1, ZBTB6). Out of these, only PPARD has been clearly associated with either of the two conditions (M-CLL), as also predicted by *diffTF*. *chromVAR*, however, associated it more with U-CLL, although the deviation value is relatively small.

HSC mouse data source (FACS sorting step)

Single-cell suspensions of mouse bone marrow were erythrolysed, enriched for Kit expression (CD117 microbeads, Miltenyi Biotec) and stained with antibodies against surface markers: Lineage (B220-PECy5 (RA3-6B2, eBioscience), CD11b-PECy5 (M1/70, eBioscience), Ter119, PECy5 (TER-119, eBioscience), CD3e-PECy5 (145-2C11, eBioscience), Gr1-PECy5 (RB6-8C5, eBioscience)), Sca1-BV421 (D7, BD biosciences), cKit-AlexaFlour 780 (2B8, eBioscience), CD150-APC (TC15-12F12.2, Biolegend), CD48-PE (HM48-1, eBioscience), CD16/32-PECy7 (93, eBioscience). The following combination of surface markers was used to define hematopoietic progenitor populations: Multipotent Progenitor (MPP) cells, Lin⁻cKit⁺Sca1⁺CD150⁻CD48⁺; Granulocyte-Monocyte Progenitor (GMP) cells, Lin⁻cKit⁺Sca1⁻CD16/32⁺. Cells were sorted on *FACSAria III* (BD Biosciences) and analyzed using the *FlowJo* software (Tree Star inc.). GMP cells from the TET2 knockout mice were isolated with the same GMP markers. All additional details such as the origin of the mouse line as well as detailed information of the knockout preparation is provided in Rasmussen et al. (2019).

HSC ATAC-Seq libraries generation

ATAC-Seq libraries were generated as described previously (Buenrostro et al., 2013; Lara-Astiaso et al., 2014), with the following modifications. Briefly, 10,000 hematopoietic progenitor cells (MPPs or GMPs) freshly isolated from individual wild-type mice were sorted into ice-cold FACS buffer (PBS + 2%FBS). The cells were pelleted using a swinging bucket centrifuge (500 x g, 10min, 4°C) with settings for low acceleration/deceleration and washed once in ice-cold PBS. The cell pellets were resuspended in 50 μ L lysis buffer (10mM Tris-HCl pH 7.4, 10mM NaCl, 3mM MgCl₂, 0.1% Igepal CA-630) by gentle pipetting and immediately centrifuged one additional time (500 x g, 10min, 4°C). The supernatant was discarded and the pellet containing released nuclei were resuspended gently in 25 μ L 1xTD buffer containing 1.25 μ L Tn5 transposase (Nextera sample preparation kit, Illumina). The transposition reaction was allowed to proceed for 45min at 37°C whereafter DNA fragments were isolated using MinElute PCR purification columns (QIAGEN) according to manufacturer's instructions.

To generate multiplex libraries, the transposed DNA were initially amplified for 5x PCR cycles using 2.5 μ L each of dual-index primers (Nextera index kit, Illumina) and 2.5 μ L PCR primer cocktail (PPC, Illumina) in a 25 μ L reaction volume of 1x KAPA HiFi hot-start ready-mix (Kapa BioSystems). The hot-start polymerase was activated prior to adding to the reaction mix by performing a brief pre-incubation step of 3min at 95°C. The amplified fragments were size-selected with AMPure XP beads (0.5X) to remove fragments larger than 600bp and an aliquot was quantified to determine the optimal PCR cycle number to obtain 1/3 of maximum fluorescence intensity (Library quantification kit, Kapa Biosystems). Finally, PCR amplification was performed using the optimal number of cycles determined for each library (max. 18 cycles in total), size-selected with AMPure XP beads (0.5X) and eluted in resuspension buffer (Illumina). The size distribution of the libraries was evaluated on Bioanalyzer (Agilent) and sequenced on NextSeq 500 (Illumina) using 75bp paired-end sequencing with an average of 25 million reads per sample.

GMP-MPP ATAC-Seq processing

We run the ATAC-Seq pipeline described as above on the four wild-type GMP and four wild-type MPP samples with default parameters using the *mm10* genome. About 30% of the reads successfully passed all the filtering criteria of the pipeline, with the majority of the reads eliminated at the "remove duplicates" and "remove mitochondrial reads" steps. The fragment length distributions also showed the typical and expected pattern, with peaks at expected lengths (e.g. mono- and dinucleotides). We used GC correction as described above to remove technical GC sequencing biases.

GMP-MPP diffTF analysis

We run *diffTF* for the GMP-MPP dataset and compared the four GMP and MPP samples using default parameters unless otherwise specified. For identifying the consensus peak set we required a minimum of two samples that need to contain the peak, which resulted in 77,678 peaks. As parameters for the peaks and TF-specific analysis, we used only the stage of the HSC differentiation (GMP or MPP) in the design formula. Due to the small number of samples and therefore also possible permutations, we used the analytical approach as described in STAR Methods. For the RNA-Seq data, we applied the same filtering as for the CLL data, see above. This resulted in 268 out of the 422 TFs to be classified as expressed.

GMP-MPP footprinting analysis

For this footprinting analysis, we used all 8 ATAC-seq samples generated for MPP and GMP cell types and downsampled all to the sample with the lowest read counts (~8.3 million) and subsequently merged them by cell type. For all TFs from the RNA-Seq data that were significantly differentially expressed in DESeq2 (adj. p value < 0.05), we then took the same set of TFBS as for *diffTF* and run the *dnase_to_javatreview.py* script from the *pyDNase* library (Piper et al., 2015) for each condition to obtain base-specific Tn5 insertions matrices for each TFBS. These values were then normalized to the amount of reads in the consensus peakset, which was calculated

using *featureCounts* from the *Subread* package with the parameters `-p -B -d 0 -D 2000 -C -Q 10 -O -s 0`. We then scaled the data by dividing by the mean average Tn5 insertions for each TF. We finally performed a PCA analysis and visualized the first two PC components, containing ~40% of the overall variance (Figure 5B). For generating the genomic background for the openness, we binned the consensus peakset with a bin size of 200 bp and randomly selected 10,000 regions for which we then ran again the *dnase_to_javatreeview.py* script followed by subsequent normalization. To generate the footprint plots for each quadrant, we divided the value of Tn5 insertions at each bp to the mean value of Tn5 insertions in the whole matrix for the specific quadrants in order to enhance the differences between them in the motif center and the surroundings.

QUANTIFICATION AND STATISTICAL ANALYSIS

All methodological details concerning the statistical tests and analyses were performed are described above in the section “Description of the *diffTF* analysis workflow.” For the used tools, particularly related to the statistical parts, are summarized in section “Tools used” - mainly R and various packages were used for significance assessment. Statistical results presented in this work, abbreviations or methodological details are either described in the main text or the figure captions in case they are shown graphically.

In summary, we used statistical techniques in *diffTF* to model confounding effects in intermediate steps (e.g., CG bias, sequencing depth differences) described in the respective sections above, minimizing the influence of known covariates for log2 fold-change estimates, taking the T statistic rather than the p value in order to not be affected too much by differences in sample sizes related to different number of binding sites across TF or generally high number of points in the background distribution. For the final significance assessment, we derived two approaches in order to account for both small number of samples and larger number of samples. For larger number of samples, a permutation-based approach that gives an empirical FDR is the preferred method, while for smaller number of samples this approach cannot be used. For the latter case, the analytical approach is suitable, and we clearly describe its differences and limitations in the sections above.

In various steps, we mitigate the effect of potential outliers and violations of statistical assumptions by normalizing data appropriately (e.g., quantile normalizing the RNA-Seq data prior to running the classification, using *DESeq2* for count normalization using a cyclic loess approach as implemented in the *normOffsets* function of *csaw* for normalizing the ATAC-Seq data), running non-parametric tests instead of their parametric equivalents (e.g., Wilcoxon rank sum test for testing foreground and background for classifying TFs into activator or repressor), summarizing data by their mean or median (e.g., for the GC binning procedure and the resulting differential TF activity measure) or capturing trends by data aggregation (e.g., over all binding sites for each TF).

The multiple testing problem is consistently addressed throughout the pipeline in different steps, whenever multiple p values are generated, and we provide both raw and adjusted p values for the final measures. The adjustment of p values is done using Benjamini-Hochberg.

Lastly, a large number of quality control plots visualize and compare the data, which enables users to judge better how well the method worked, whether assumptions are met or whether data show any types of irregularities.

DATA AND CODE AVAILABILITY

The *diffTF* pipeline as well as a detailed documentation are publicly available as a *Snakemake* workflow through <https://difftf.readthedocs.io/en/latest/>, from which the Github repository at <https://git.embl.de/grp-zaugg/diffTF> is also linked. The CLL dataset treated with Ibrutinib supporting the current study has not yet been deposited in a public repository because it is part of ongoing work, but is available from the lead author upon request.